# 目　录

# 第一章 绪 论

## §1.1 数字图像处理技术的概况

### 一、图像处理技术的发展与应用

早在本世纪20年代，人们利用巴特兰(Bartlane)电缆图片传输系统，经过大西洋传送了第一幅数字图像，它使传输的时间从一个多星期减少到了三小时，使人们感受到数字图像传输的威力。它的传输方法，首先是对图像进行编码，然后在接收端用一台电报打印机利用字符模拟中间色调把图像还原出来，这是个初步尝试。为了对图像的灰度、色调和清晰度进行改善，人们采用各种方法对图像的传输、打印和恢复等技术进行改进，这种努力一直延续到此后的40年。直到大型计算机出现后，人们才开始用计算机来改善图像。在1964年，美国喷射推进实验室(JPL)进行了太空探测工作，当时用计算机来处理测距器7号发回的月球图片，以校正飞船上电视摄象机中各种不同形式的固有的图像畸变，这些技术都是图像增强和复原的基础。同时他们成功地用计算机绘制出月球表面的地图。随后在1965年又对徘徊者8号发回的几万张照片进行较为复杂的数字图像处理，使图像的质量进一步提高。JPL的工作引起了世界许多有关方面的注意，JPL也更加重视数字图像技术的研究，投入了更大的力量，并取得了许多非凡的成果。与此同时，JPL以及世界各国有关部门已把数字图像处理技术从空间技术开发到生物医学、X射线图像增强、光学显微图像分析、遥感图像分析、粒子物理、地质勘探、人工智能和工业检测等等方面。

数字图像处理技术在近20多年的时间里，迅速地发展成为一门独立的有强大生命力的学科，下面仅就几个方面的某些应用举些例子。

1. *遥感技术*

遥感技术可以是飞机遥感和卫星遥感技术，从前，许多国家每天派出很多侦察飞机对地球上感兴趣的地区进行大量的空中摄影。对由此得来的照片进行判读分析需要雇用几千人，而现在改用配备有高级计算机的图像处理系统来判读分析，既节省人力，又加快速度，还可以从照片中提取人工所不能发现的大量的有用情报。从遥感卫星所获得的地球资源图片由于各种原因，图像质量总不是很好，如果仍采用简单的直观判读如此昂贵代价所获取的图像是不合算的，因此必须采用图像处理技术。如LANDSAT系列陆地卫星，采用多波段扫描器(MSS)，在900公里高空，对地球每一个地区以18天为一周期进行扫描成象，其图像分辨率大致相当地面上十几米或一百米左右。这些图像无论在成象、存贮、传输过程中，还是在判读分析中，都必须采用很多的数字图像处理方法。目前遥感技术，尤其是卫星遥感，已经在资源调查、灾害监测、农业规划、城市规划、环境保护等方面取得了很大的应用效果。我国也在以上诸方面的实际应用中取得了良好的成果，对我国国民经济的发展起到了相当大的作用。

2. 医用图像处理

医学上不管是基础科学还是临床应用，都是图像处理种类极多的领域，例如对生物医学的显微图像的处理分析方面，如红白细胞和细菌，染色体分析，另外像胸部X线照片的鉴别，眼底照片的分析，以及超声波图像的分析等等都是医疗辅助诊断的有力工具，目前这类应用已经发展到专用的软件和硬件设备，最普遍使用的是计算机层析成象，亦称为CT技术。它是由英国的Hounsfield和美国的Cormack发明的。通过CT，可以获取人体剖面图，使得肌体病变特别是肿瘤诊断起到了革命性的变化，两位发明者因此获得1979年诺贝尔医学奖。近年来又出现的核磁共振CT，使人体免受各种硬射线的伤害，并且图像更为清晰。图像处理技术在医学上的应用正在进一步的发展。

3. 工业领域中的应用

在工业领域中的应用一般有以下几方面：工业产品的无损探伤，表面和外观的自动检查和识别，装配和生产线的自动化，弹性力学照片的应力分析，流体力学图片的阻力和升力分析。其中最值得注意的是“计算机视觉”，采用摄影和输入二维图像的机器人，可以确定物体的位置、方向、属性以及其它状态等，它不但可以完成普通的材料搬运，产品集装，部件装配，生产过程自动监控，还可以在人不宜进入的环境里进行喷漆、焊接、自动检测等等。现在已发展到具备视觉、听觉和触觉反馈的智能机器人。

4. 军事公安方面

主要应用是：各种侦察照片的判读，运动目标的图像自动跟踪技术，例如目前电视跟踪技术已经装备到导弹和军舰上，并在实践和演习中取得很好的效果。另外还有公安业务图片的判读分析，如指纹识别，不完整图片的复原等等，在公安中的跟踪、窃视、交通监控、事故分析中都已经用到了图像处理的技术。

5. 文化艺术方面

在文化艺术方面有电视画面的数字编辑，动画片的制作，服装的花纹设计、制作。文物资料照片的复制和修复。在体育方面，运动员的训练，动作分析和评分等等。

随着计算机技术的日益发展，图像处理技术的日益完备，图像处理的应用范围将越加深入和广泛。

## 二、数字图像处理技术的特点

在计算机处理出现以前，图像处理都是光学照像处理和视频信号处理等模拟处理，把数字图像与其它模拟方法在以下各方面比较如下：

| 方式 | 处理速度 | 灵活性 | 精度 | 调整 | 再现性 | 其它 |
| --- | --- | --- | --- | --- | --- | --- |
| 光学 | √ | ○ | ○ | × | ○ | ×现象 |
| 照片 | √ | × | × | × | × | ×现象 |
| 录像 | √ | ○ | × | ○ | ○ | |
| 数字 | × | √ | √ | √ | √ | ×内存 |

从上面的列表可以看出，除了处理速度和内存要求大以外，数字图像处理技术在灵活性、精度、调整和再现性方面都是卓越的，它具有用程序能自由地进行各种处理，并且能达到较高的精度。这与模拟处理中，要提高一个数量级的精度，就必须对装置进行大幅度改进相比确实

为一优点。另外，由于半导体技术的不断进步，实际上正在开发出普遍使用的微处理器的图像处理专用高速处理器，以 IC 存贮器为基础的图像显示也达到可行的程度，这些都进一步加快了数字图像处理技术的发展和实用化。

为了用计算机处理图像，必须把图像作为数值来表示，数字图像就是二维平面上的灰度分布。数字图像信息有以下的特点。

(1)信息量很大。例如一帧电视图像取 512×512 个象素组成，如其灰度级用 8 比特的二进位制来表示，则有 $2^8=256$ 个灰度级，那么一帧图像的信息量即为 512×512×8=2 097 152 比特。对这样大信息量的图像进行处理，必须要有计算机才能胜任，而且计算机的内存量要大。

(2)数字图像占用的频带较宽。与语言信息相比，占用的频带要大几个数量级。如电视图像的带宽为 5～6 MHz，而语言带宽仅为 4 kHz 左右。频带愈宽，技术实现的难度就大，成本亦高，为此对频带压缩技术提出了较高的要求。

(3)数字图像中各个象素是不独立的，其相关性很大。例如在电视画面中，同一行中相邻两个象素或相邻两行间的象素，其相关系数可达 0.9，而相邻两帧之间的相关性比帧内相关性还要大一些。因此图像信息压缩的潜力很大。

(4)处理后的数字图像是需要给人观察和评价的，因此受人的因素影响较大。由于人的视觉系统很复杂，受环境条件、视觉性能、人的主观意识的影响很大，因此要求系统与人的良好的配合，这还是一个很大的研究课题。

## §1.2 视觉的原理与模型

### 一、视觉的原理

光线照在物体上其透射或反射光的分布就是“图”。而人的视觉系统对图的接收在大脑中形成的印象或认识就是“象”。前者是客观存在，后者为人的感觉，图像就是二者的结合。下面首先谈谈人眼的构造。

在图 1-1 中画出了人眼的横断面，眼睛的形状为一圆球，即平均直径约为 20 mm，它由三层薄膜包着，即角膜和巩膜外壳，脉络膜和视网膜。角膜是一种硬而透明的组织，它盖着眼睛的前表面，巩膜与角膜连在一起，巩膜是一层包围着眼球剩余部份的不透明的膜。脉络膜位于巩膜的里边，这层膜包含有血管网；它是眼睛的重要滋养源，脉络膜外壳着色很重，因此有助于减少进入眼内的外来光和眼球内的回射。在脉络膜的最前面被分为睫状体和虹膜。虹膜的收缩和扩张控制着允许进入眼内的光量。虹膜的中间开口处(瞳孔)的直径是可变的，约由 2 mm 变到 8 mm。虹膜的前部含有眼睛的明显的色素，而后部则含有黑色素。

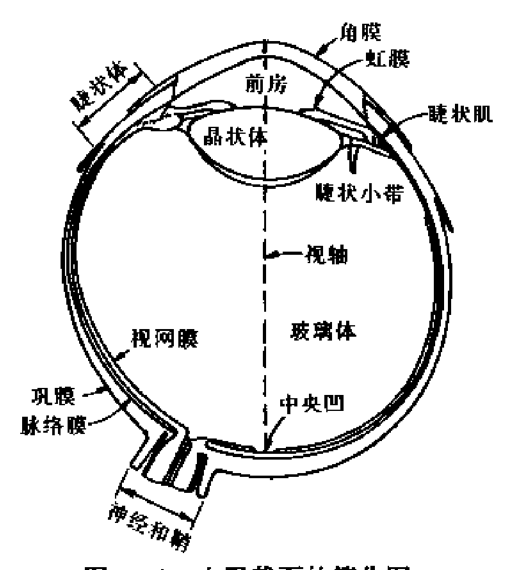


图 1-1 人眼截面的简化图

眼睛最里层的膜是视网膜，它布满在整个后部的内壁上，当眼球被适当的聚集，从眼睛的

外部物体来的光就在视网膜上成象。整个视网膜表面上分布的分离的光接收器造成了图案视觉。这种光接收器可分为两类：锥状体和杆状体。每只眼睛中锥状体的数目在 600 万到 700 万之间。它们主要位于视网膜的中间部份，叫做中央凹，它对颜色很敏感，人们用这些锥状体能充分地识别图像的细节，因为每个锥状体都被接到其本身的神经的一端。控制眼睛的肌肉使眼球转动，从而使人所感兴趣的物体的像落在视网膜的中央凹上，锥状视觉又叫白昼视觉。

杆状体数目更多，约有 7500 万到 15000 万个，分布在视网膜表面上，因为分布面积较大并且几个杆状体接到同一根神经的末端上，因而使接收器能够识别的细节的量减少了，杆状体用来给出视野中一般的总的图像。它没有色彩的感觉而对低照明度的景物则较敏感。例如，在白天呈现鲜明颜色的物体，在月光之下却没有颜色，这是因为只有杆状体受到了刺激，而杆状体没有色彩的感觉。杆状视觉又叫夜视觉。

眼睛中的晶状体与普通的光学透镜之间的主要差别在于前者的适应性强，如图 1-1 所画，晶状体前面的曲率半径大于后表面的曲率半径。晶状体的形状由睫状体的韧带的张力来控制。为了对远方的物体聚焦，控制用的肌肉就使晶状体变得较扁平。同样，为使眼睛近处的物体得到聚集，肌肉就使晶状体变得较厚。

当晶状体的折射能力由最小变到最大时，晶状体的聚集中心与视网膜之间的距离约由 17 mm缩小到 14 mm。当眼睛聚焦到远于 3 m 的物体时，晶状体的折射能力最弱，当聚集到非常近的物体时，其折射能力最强。利用这一数据，将易于计算出任何物体在视网膜上形成图像的大小。例如图 1-2 中，观测者看一个距离为 100 m，高为 15 m 的房子，设 $x$ 为视网膜上形成的图像的大小，单位为 mm，由图 1-2 的几何形状来看，则有 $15/100=x/17$，或$x=2.55$mm。正如前面所指出的，视网膜图像主要反映在中央凹的面积上。然后由光接收器的相对刺激作用产生感觉。这样，感觉把辐射来的能量转变为电脉冲，最后由脑子判别出来。

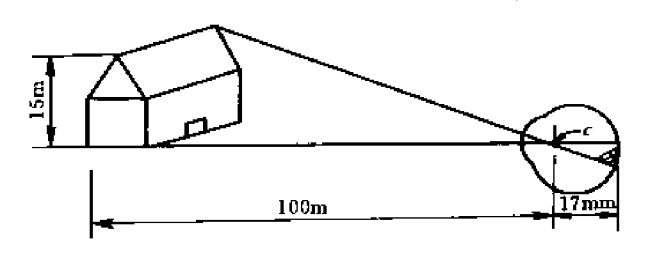


图 1-2　用眼睛观察房子的光学表示法，c 点为晶状体的光学中心

由于数字图像是作为许多分离的亮点显示出来的，因此眼睛对于不同亮度之间的鉴别能力，是图像处理结果中所要考虑的一个重要方面，人的视觉系统能够适应光强度的级别的范围是很宽的，由夜视阈值到强闪光之间光强度的级别约为 $10^{10}$级，相当多的实验证据表明主观亮度是进入眼内的光强度的对数函数。这一特性画于图 1-3，它表示光强度与主观亮度之间的关系。长的黑线代表人的视觉系统所能适应的光强度的范围。昼视觉范围为 $10^{6}$，由夜视觉到昼视觉是逐渐过渡的，过渡的范围大致由 0.001 到 0.1 毫朗伯①(在对数标度中为 −3 mL 到 −1 mL)在图中画出了在这一范围内的适应曲线。

解释画在图 1-3 中的特殊动态范围的关键在于人的视觉绝对不能同时在这一范围工作，说得更确切一点，它是利用改变其整个灵敏度来完成这一大变动的，这就是所谓亮度适应现象。当与整个适应范围相比时，能同时鉴别光强度级的总范围是很小的。对于任何一组给定的条件，视觉系统的当前的灵敏度级叫做亮度适应级，例如它相当于图 1-3 中的亮度 $B_a$，短的

① 1 朗伯 $=1\ \mathrm{lm/cm^2}$

交叉曲线表示当眼睛适应于这一强度级时，人眼能感觉的主观亮度的范围。必须注意这一范围是有一定限制的，在 $B_b$ 处和 $B_b$ 以下时，所有的刺激都是作为不可分辨的黑来理解的。

眼睛的对比灵敏度可用下述方法测量出来，使观测者处于亮度为 B 的均匀光场中，中部有一个具有鲜明边界的园形靶，其亮度为 $B+\Delta B$，如图 1－4(a)所示。$\Delta B$ 由零值一直升到刚好被觉察到的值为止。这个刚能觉察到的亮度差值 $\Delta B$ 是作为 $B$ 的函数而被量测的。$\Delta B/B$ 称为韦伯比，在非常宽的亮度范围内，它近似为一常数，约为 2%，如图 1－4(b)所示。这一现象导致下述看法，即人眼具有比人造图像系统宽得多的动态范围，如果利用图 1－5(a)的图案可得到更加适用的结果，再一次测出 $\Delta B/B$，但这里的 $B_0$（周围亮度）是一个参数，图 1－5(b)中画出了结果。动态范围在以周围亮度为中心的 2.2 单位内变动，如果背景亮度调节正确的话，那么这一结果可以与用电学图像系统所能达到的结果相比较。真正引人注意的特性是眼睛自身适应的难易强度和快慢（在视网膜的不同部份是不同的），而不是它的整个动态范围。

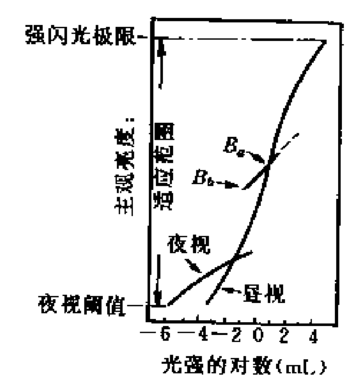



图 1－3　表示特殊适应能力的主观亮度感觉的范围

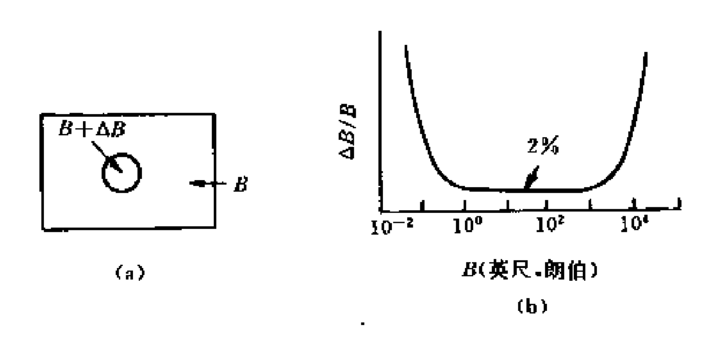



图 1－4　具有恒定背景的对比灵敏度

在复杂图像的情况下，视觉系统不会适应于单一光强度级。相反，它适应于一平均光强度级，这一平均光强度级取决于图像的性质。当眼睛在景物周围徘徊时，瞬时适应能力围绕其平均值上下波动。在图像中任何一点或小面积上，韦伯比一般要比在实验环境中得到的值大得多，这是因为在背景中缺乏鲜明的限制边界和强度变化。其结果，眼睛只能在一个复杂的图像中的任何一点上的十多个或廿多个亮度级附近进行检测。但这并不意味着图像只需要用廿多个光强度级进行显示就能获得令人满意的视觉效果。上面的窄的鉴别范围跟踪着适应力而变化，因为适应能力需要变化，以便当眼睛沿着景物移动时，适应不同的光强度级。这允许有大得多的全光强度鉴别的范围。对于一大类图像类型来说，为了使眼睛得到适当舒服的显示，一般需要大于 100 光强度级的范围。

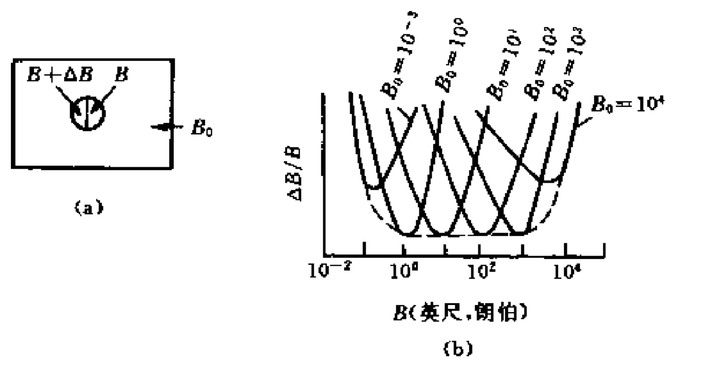



图 1－5　具有可变背景的对比灵敏度

人眼能感觉到的一个区域的高密度取决于许多因素，而不是简单地取决于由该区幅射出的光，根据图像处理的用途，有关亮度感觉的最有价值的现象是人的视觉系统的响应在不同光强度区的边界周围引起的“过量调整”。这一过量调整的结果会造成某个恒定光强度区域的出现，如同这些区域原来就具有可变的亮度一样。例如，在图 1－6(a)中所画的图像是根据照片

下面的光强度分布而形成的。虽然光强度变化是完全圆滑的，但人的眼睛所感受到的是 $B$ 区中较亮的带和 $D$ 区中较暗的带。这些带叫做马赫带。马赫带效应的更加明显的例子画于图 1－6(b)中。正如光强度分布图所指示的那样，照片中的每一条带都是由一固定光强度造成的，然而对于眼睛来说，图像中的亮度图案，特别在边界的周围，出现了强度的变化。

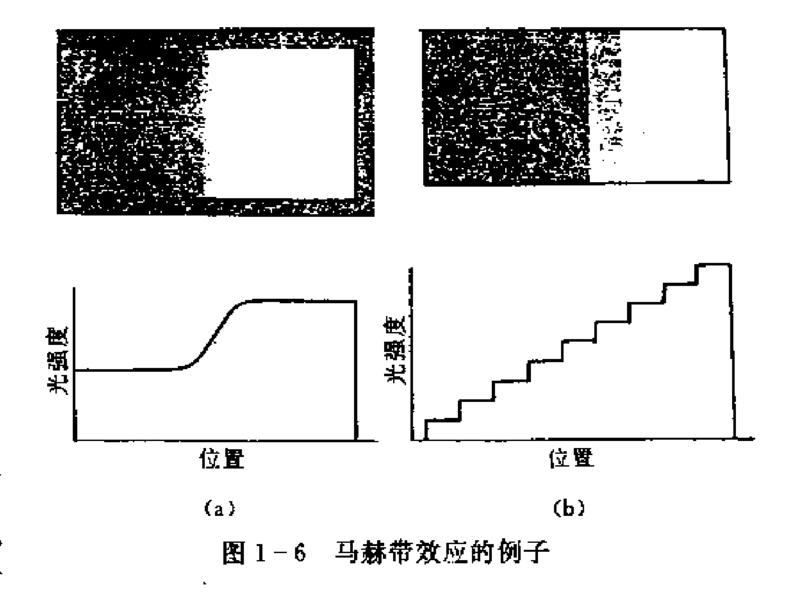


图 1－6　马赫带效应的例子

## 二、视觉模型

将视觉系统的功能抽象化为简单的模型。并以此模型为基础来对视觉系统的功能进行研究或加以模仿是很吸引人的。建立模型首先要提出假说，根据假说建立模型，再根据试验修改模型，反复进行使之逐渐完善合理。建立模型可以有电子线路模型以及化学模型等，对我们来说，一般是对电子线路模型感兴趣，这样可以把视觉系统的一些优越性引入图像通信和信息处理系统加以研究和应用。

下面简单地介绍一下神经元模型，黑白视觉模型和彩色视觉模型

### （一）神经元模型

从信息处理观点出发，在神经元所具有的各种机能中最重要的是，在突触处许多输入在空间和时间上进行加权的性质，以及细胞的阈值作用。有两种神经元模型，一种是针对研究空间加权特性的不计时间特性的神经元模型，另一种是考虑到时间特性的空间加权特性的模型。现对后一种略加介绍。

*1. 神经元信息的产生及传递特性*

神经元由细胞体及其轴突（神经纤维）和树突（细胞体粗短突起）组成，轴突末端称为末梢，一个神经元末梢与另一个神经元细胞体或树突相接触形成突触，神经元之间的信息传递都是通过突触进行的，见图 1－7[1-1]。平常细胞膜内外保持一个约 70 mV 的电位差，当细胞兴奋时就会发出幅度为 100 mV，宽度为 1 ms 左右的脉冲电位，其脉冲的频率随细胞兴奋程度而变。这一电位经突触传给另一个神经元，产生所谓突触后电位简称 PSP(Post Synaptic Potential)。把能够产生的 PSP 的突触称为兴奋性突触，它产生的兴奋性突触后电位，简称 EPSP，与此相反地还有抑制性突触产生的抑制性后电位 IPSP。

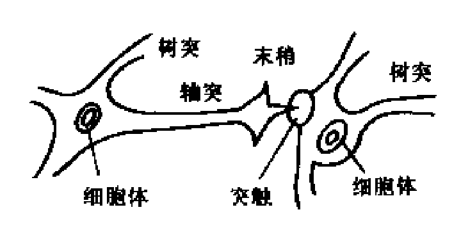


图 1－7　神经元之间突触联系

在神经元内部之间的传输信息（兴奋）是有速度的，约每秒几米到几十米，因此神经元有时间效应。

轴突有保持某一固有脉冲的性质，比较小的波形在传递中可以增幅，比较大的波形在传递中可以衰弱。对脉冲宽度有同样的性质，因此有相似于自动增益控制作用。不管输入波形有什么变化，输出波形是恒定的，而且对同类波形可以恒定速度传送。

但轴突传送的信号，对于它的脉冲幅度要有一定的阈值，比阈值小的信号在传送过程中就消失了，这种阈值作用相似于静化噪声的作用。神经元的阈值变化是这样的：在发放之后有一段时间约为 1 ms，无论给予多么强的刺激都不再发放，称为绝对不应期。稍后阈值慢慢下降并恢复到原来状态，这段慢慢下降期，即阈值比正常值高又难于引起反应的期间称为相对不应期。一般来讲，信息在轴突中的传送是单方向的，而突触传送也有单向也有双向的。

2. *考虑时间特性的神经元脉冲模型*

在考虑时间特性的模型中，有些是和实际神经元同样产生脉冲输出的模型，也有电子线路模型[1-2]。其主要功能框图如 1－8 所示。

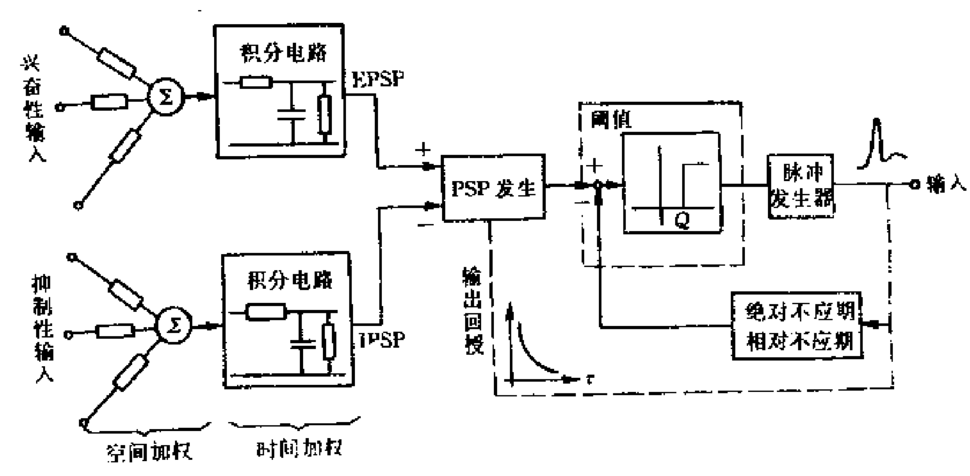



图 1－8　脉冲输出神经元模型

模型的输入输出都是脉冲序列，由 EPSP 和 IPSP 的干涉（如两者的差或和）来决定 PSP，当 PSP 超过一定阈值 $Q$ 时，产生脉冲输出，脉冲发生后进入绝对不应期，脉冲放电暂时被抑制，接着进入了相对不应期，随着阈值恢复正常，下一个脉冲立即发生，PSP 越大，脉冲重复频率越高。

**（二）黑白视觉模型**

研究下列线性光学系统如图 1－9 所示。

设 $\varphi_i(\omega_x,\omega_y)$ 为输入图像 $I_i(x,y)$ 的傅里叶变换，$\varphi_o(\omega_x,\omega_y)$ 为输出图象 $I_o(x,y)$ 的傅里叶变换。根据线性系统理论可得：

$$\varphi_o(\omega_x,\omega_y)=H(\omega_x,\omega_y)\cdot\varphi_i(\omega_x,\omega_y)$$

在许多情况下，人们只对相对输入图像强度和振幅变化的输出图像强度感兴趣，因此由下式描述

$$H(\omega_x,\omega_y)=\frac{\varphi_o(\omega_x,\omega_y)}{\varphi_i(\omega_x,\omega_y)}$$

式中　$H(\omega_x,\omega_y)$ 称为该光学系统的调制传递函数 MTF。

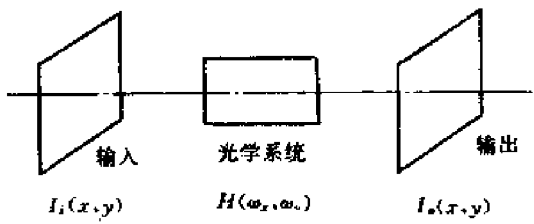



图 1－9　线性光学系统

把光学系统的概念用到人的视觉系统方面已经作出了很多努力，从实验结果可以看出，MTF 测试结果与输入对比度的大小有关，而且如果当输入正弦光栅相对于眼球光轴进行旋转时，则测得的 MTF 形状也有某些改变。于是可以推出以下结论：人的视觉系统由实验测出的 MTF 是非线性的和各向异性的（即旋转可变的）。另外，根据一些人的实际结果[1-3]可以认为，眼睛对光强度的非线性响应呈对数型，并且发生在视觉系统的开始附近（亦就是视觉信号在锥

状及杆状细胞空间上发生相互作用之前)。由此得出,人眼黑白视觉的简单对数模型如图 1-10 所示。

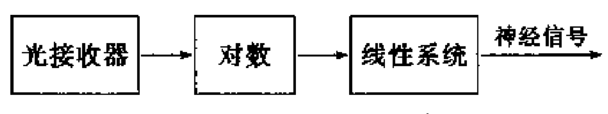


图 1-10 黑白视觉对数模型

下面介绍一个表示黑白视觉扩展的模型,如图 1-11 所示。在模型中把对不同波长有不同灵敏度的感受器输出馈送到代表眼睛光学性能的低通线性系统 $H_1(\omega_x,\omega_y)$,接着是代表杆状和锥状视细胞非线性强度响应的一般黑白非线性网络,下面就是代表侧抑制过程的具有带通性能的线性系统 $H_2(\omega_x,\omega_y)$,最后是具有时间滤波效应的线性系统 $H_3(\omega_x,\omega_y)$。

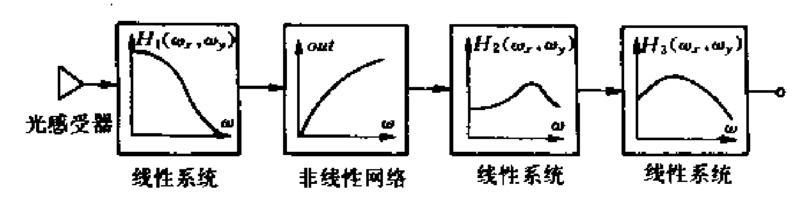


图 1-11 黑白视觉扩展模型

**(三)彩色视觉模型**

根据三色假说为基础来研究彩色视觉模型,图 1-12 是 *Frei* 建议的彩色模型。在该模型中,$e_1$,$e_2$,$e_3$ 代表视网膜三个具有 $S_1(\lambda)$,$S_2(\lambda)$,$S_3(\lambda)$谱灵敏度的感受器,其输出分别为

$$e_1=c(\lambda)S_1(\lambda)d\lambda$$

$$e_2=c(\lambda)S_2(\lambda)d\lambda$$

$$e_3=c(\lambda)S_3(\lambda)d\lambda$$

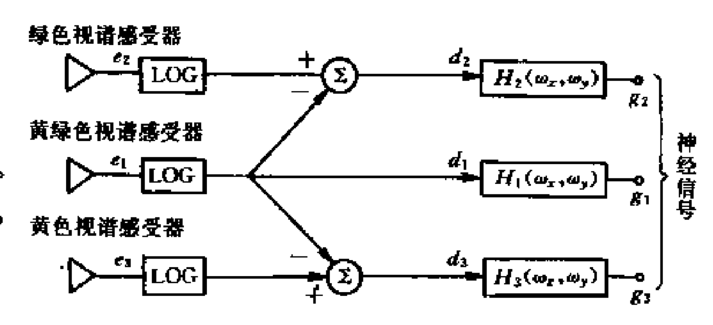


图 1-12 彩色视觉模型

式中 $c(\lambda)$为入射光源的谱线分布函数。$e_1$,$e_2$,$e_3$ 经对数传递后并组合为 $d_1$,$d_2$,$d_3$ 输出

$$d_1=\lg(e_1)$$

$$d_2=\lg(e_2)-\lg(e_1)=\lg\frac{e_2}{e_1}$$

$$d_3=\lg(e_3)-\lg(e_1)=\lg\frac{e_3}{e_1}$$

最后,信号 $d_1$,$d_2$,$d_3$ 分别经传递函数 $H_1(\omega_x,\omega_y)$,$H_2(\omega_x,\omega_y)$,$H_3(\omega_x,\omega_y)$的线性系统输出 $g_1$,$g_2$,$g_3$。由这些信号提供大脑感受彩色的基础。

信号 $d_2$ 和 $d_3$ 与彩色光的色度有关,而 $d_1$ 则正比于它的亮度。这个模型相当准确地预测许多彩色现象,也能满足色度学的基本定律。例如,彩色光的谱能量被改变为一个常数,则光线的色调和饱和度正如色度坐标定量描述那样,在一个宽的动态范围内是不会变的,从以上的公式可以看出 $d_2$ 和 $d_3$ 彩色信号是不会变的,而亮度信号 $d_1$ 则按对数规律增加。

与图 1-11 所示的黑白视觉扩展模型类似,对数的彩色视觉模型很容易被扩展,考虑到眼睛的响应,只要在每一种锥状视细胞感受器后插入一个线性传递函数即可。同样,可用一般的非线性网络代替对数传递函数网络。应当注意,感受器的相加和传递函数的运算次序可以互换而不影响其输出,因为两者都是线性运算,图 1-13 为彩色视觉扩展模型。

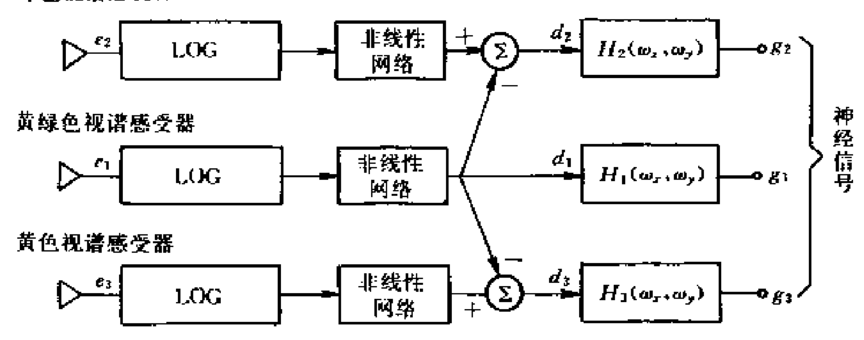



图 1－13　彩色视觉扩展模型

由 Vonkrieo[1-4]提出的关于彩色适应最简单的模型，就是在图 1－13 中的锥状细胞感受器和第一个线性系统之间插入自动增益控制，其增益为

$$a_i=[w(\lambda)S_i(\lambda)d\lambda]^{-1} \qquad i=1,2,3$$

被这样调整后，当观察参考白色及谱能分布为 $w(\lambda)$图像时，修正后的锥细胞响应为 1。Vonkrior 的模型是很吸引人的，因为它在定性上是合理的，而且简单。但实际彩色测试证明，该模型并不能完全预测彩色适应效应。Wallio 提出彩色适应部分是由网膜在神经抑制机构中所产生的，该机构使图像中变化分量慢慢地线性衰减。这可将图 1－13 中网膜后传递函数 $H_1(w_x,w_y)$的低空间频率衰减用来模型化。毫无疑问，视神经和网膜后机构都可能产生彩色适应效应，为此更精确的模型尚须进一步研究。

## §1.3　图像的数字化

我们已经知道，计算机所能处理的信息必须是数字信号，而我们得到的照片、图纸或景物等信息都是连续信号，为此必须将此连续信号进行抽样和量化，即进行数字化处理。设连续图像 $f(x,y)$经过等间隔抽样以后，可以用一个离散量组成的矩阵来表示

$$f(x,y)\approx\begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,n-1) \\ f(1,0) & f(1,1) & \cdots & f(1,n-1) \\ f(n-1,0) & \cdots & \cdots & f(n-1,n-1) \end{bmatrix}$$

矩阵中的每一个元素称作象元、象素或图像元素。而 $f(x,y)$代表$(x,y)$点的灰度值，即亮度值。以上数字化有以下几点补充说明一下。

(1)由于 $f(x,y)$代表该点图像的光强度，而光是能量的一种形式，故 $f(x,y)$必须大于零，且为有限值，即

$$0<f(x,y)<\infty$$

(2)以上数字化抽样是按正方形点阵取样的，除此之外还有三角形点阵，正六角形点阵取样。

(3)以上是用 $f(x,y)$的数值来表示$(x,y)$位置点上灰度级值的大小，亦即只反映了黑白灰度的关系，如果是一幅彩色图像，各点的数值还应当反映出色彩的变化，即可用 $f(x,y,\lambda)$表示，其中 $\lambda$ 为波长。如果是一幅活动的彩色图像，还应是时间 $t$ 的函数，即可表示为 $f(x,y,\lambda,t)$。

(4)以上数字化后的矩阵为 $N\times N$ 的方阵，对各象素允许的最大灰度级数都要作出决定，一般来说，无论是阵列大小 $N$ 和象素的最大灰度级数 $G$ 都取为 2 的整次幂，即

$$N = 2^n, G = 2^m$$

而对 $N \times N$ 的象素，具有 $G$ 级灰度级时，则存贮此数字化图像所需的位数为 $b$，它的单位为比特，即

$$b = N \times N \times m$$

例如，灰度级 $G=64$ 级(6 bit)的 128×128 图像需要 98304 个存贮位，图像的清晰度(即可辨别的细节的程度)主要取决于 $N$ 和 $m$，这些参量越大，数字阵列对于原来的图像的近似就越好，但是存贮量以及由此而引起的计算量也作为 $N$ 和 $m$ 的函数而很快地增加。

对 $N$ 与 $m$ 的选择，应根据图像的性质与处理的目的来决定，由于微型机的普及与发展，$m$ 多采用 8 bit，即 256 个灰度级。

必须指出，当 $N$ 和 $m$ 增加时，并不是所有图像的质量都会提高，在个别情况下，图像质量随 $m$ 的减少而得到改善。这是因为 $m$ 减少会使图像的对比度增加。但往往 $m$ 太小会使图像出现"假轮廓"，为此一般 $m$ 至少取为 6，即 64 个灰度级。有时也可以采用非均匀抽样和非均匀量化来减少计算量，但是这样做容易使问题复杂化，因此很少被采用。

最后举出几个实际处理的象素数的典型例子：

汉字：32×32～64×64 象素。如果不仅要区别作为文字的图形，还要处理细小部份的形状时，则需要更多的象素数。

显微镜像：256×256～512×512 象素。

TV 图像：500～700×480 象素。

卫星图像：LANDSAT(单波段)3240×2340 象素。

SAR(合成孔径雷达)：8000×8000 象素左右。

CRT 显示器：作为显示画面，经常采用 512×512～1024×1024 象素的显示器，亦可以采取更高象素的显示器。

## 习　题

1—1　试举例说明数字图像处理的应用。尽量结合自己的实际工作和生活。

1—2　试说明人类眼睛的构造，以及眼睛中成像的过程。

1—3　试说明三种视觉模型，画出黑白视觉扩展模型，并略加说明。

1—4　试叙述图像矩阵、$f(x,y)$，$N$ 和 $m$ 的意义？试说明对文字、显微、TV、卫星、CRT 等图像取用象素的数值。

## 参 考 文 献


[1—1] Winte P. A. Transform Coding Transmission. IEEE Speetrum, Vol. 2, No. 12, 1995.

[1—2] Pottala E. W. ,Colburn T. R. ,Humphrery D. R. ,A Dendritic Compartment Model Nerun IEEE
Trans,Biomed,Eng. BME-20(2)March 1993 P132～139

[1—3] Pratt W. K. Digital Image Processing,1978

[1—4] Von Kries J. Die Gesichtsempfindungen Nagel's Handbuch d,physiology d. Menchen, Vol. 3,1904,211

[1—5] 冈萨雷斯 R.C. 数字图像处理. 李叔梁等译，北京：北京科学出版社，1983

# 第二章 图像的变换

在信号与系统课程中，我们已经把信号处理方法主要分成两大部分，一是时域分析法，二是频域分析法，在图像处理技术中同样存在以上两种方法。把信号或图像变换到频率域可以从另外一个角度来分析信号的特性，便于我们更准确地处理它，往往利用频率域的特性分析和处理图像将更加实用一些。

众多的图像变换方法不断涌现，从古老的傅里叶变换，发展到余弦变换以至最近风靡世界的子波变换。从沃尔什变换到哈达玛变换到哈特林变换等等。所有的变换虽然名称各不相同，但有一点是共同的，也就是每一个变换都存在自己的正交函数集，正是由于各种正交函数集的不同而引入不同变换。正如表示空间的一个矢量可以用不同的坐标系一样，变换的途径虽然不同，但它们都是空间域图像 $f(x,y)$的变换域表示式。

本章将从傅里叶变换入手，虽然它已经古老，但是它的理论完善，应用程序多。如把一种变换掌握透彻，对其它类的变换可以起到触类旁通的作用。本章还将介绍余弦变换，哈达玛变换，沃尔什变换，并对子波变换给予简单的介绍。

## §2.1 傅里叶变换

当信号 $f(x)$满足狭里赫利条件时，$f(x)$的傅里叶变换式定义为

$$F(u)=\int_{-\infty}^{\infty}f(x)e^{-j2\pi ux}dx \qquad (2.1.1)$$

其反变换式为

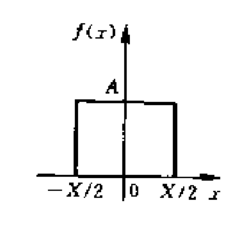
$$f(x)=\int_{-\infty}^{\infty}F(u)e^{j2\pi ux}du \qquad (2.1.2)$$

在此，$f(x)$必须满足只有有限个间断点、有限个极值和绝对可积的条件，并且 $F(u)$也是可积的。实际上以上条件一般情况下总是可以满足的。$f(x)$一般是实函数，而 $F(u)$是一个复函数，它由实部和虚部组成

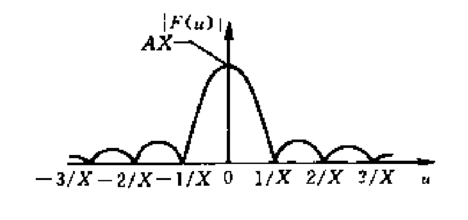
$$F(u)=R(u)+jI(u)=|F(u)|e^{j\phi(u)} \qquad (2.1.3)$$

$$|F(u)|=\sqrt{R^2(u)+I^2(u)}$$

$$\phi(u)=\mathrm{arctg}\,\frac{I(u)}{R(u)}$$




(a)




(b)

图 2-1

以 $F(u)\sim u$ 画出的频谱为幅谱，而 $\phi(u)\sim u$ 为相谱。

$E(u)=F(u)^2=R^2(u)+I^2(u)$则称为能量谱。

例 1：求一维方波信号的傅里叶变换

$$f(x)=\begin{cases}A & |x|<\frac{\tau}{2}\\ 0 & |x|>\frac{\tau}{2}\end{cases}$$

$$F(u)=\int_{-\infty}^{\infty}f(x)\mathrm{e}^{-\mathrm{j}2\pi ux}\mathrm{d}x=\int_{-\tau/2}^{\tau/2}A\mathrm{e}^{-\mathrm{j}2\pi ux}\mathrm{d}x$$

$$=\frac{A}{\mathrm{j}2\pi u}(\mathrm{e}^{\mathrm{j}2\pi u\frac{\tau}{2}}-\mathrm{e}^{-\mathrm{j}2\pi u\frac{\tau}{2}})=A\tau\frac{\sin(\pi u\tau)}{\pi u\tau}=A\tau\mathrm{sinc}(\pi u\tau)$$

从以上一维傅里叶变换可以容易地推广到二维傅里叶变换，如果 $f(x,y)$ 是满足狄里赫利条件，那么 $f(x,y)$ 的二维傅里叶变换 $F(u,v)$ 必然存在

$$F(u,v)=\iint_{-\infty}^{\infty}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y$$

$$f(x,y)=\iint_{-\infty}^{\infty}F(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux+vy)}\mathrm{d}u\mathrm{d}v \tag{2.1.4}$$

式中 $u,v$ 是频率变量。与一维傅里叶变换一样，二维函数的傅里叶谱、相位谱和能量谱分别由下列关系式给出：

$$|F(u,v)|=[R^2(u,v)+I^2(u,v)]^{1/2}$$

$$\varphi(u,v)=\mathrm{arctg}\frac{I(u,v)}{R(u,v)} \tag{2.1.5}$$

$$E(u,v)=R^2(u,v)+I^2(u,v)$$

对应于二维方波信号

$$f(x,y)=\begin{cases}A & 0\leqslant x\leqslant X,0\leqslant y\leqslant Y\\ 0 & \text{其它}\end{cases}$$

傅里叶谱为

$$F(u,v)=\iint_{-\infty}^{\infty}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y$$

$$=A\int_0^X\mathrm{e}^{-\mathrm{j}2\pi ux}\mathrm{d}x\int_0^Y\mathrm{e}^{-\mathrm{j}2\pi vy}\mathrm{d}y$$

$$=A\left.\frac{\mathrm{e}^{-\mathrm{j}2\pi ux}}{-\mathrm{j}2\pi u}\right|_0^X\left.\frac{\mathrm{e}^{-\mathrm{j}2\pi vy}}{-\mathrm{j}2\pi v}\right|_0^Y$$

$$=AXY\cdot\frac{\sin\pi uX}{\pi uX}\mathrm{e}^{-\mathrm{j}\pi uX}\cdot\frac{\sin\pi vY}{\pi vY}\mathrm{e}^{-\mathrm{j}\pi vY} \tag{2.1.6}$$

其幅度谱为

$$|F(u,v)|=AXY\left|\frac{\sin\pi uX}{\pi uX}\right|\left|\frac{\sin\pi vY}{\pi vY}\right| \tag{2.1.7}$$

相应的图形如图 2-2 所示，其中(a)为原函数，(b)为 $f(x,y)$ 的傅里叶谱 $F(u,v)$，(c)为(b)的以强度函数显示的图形。

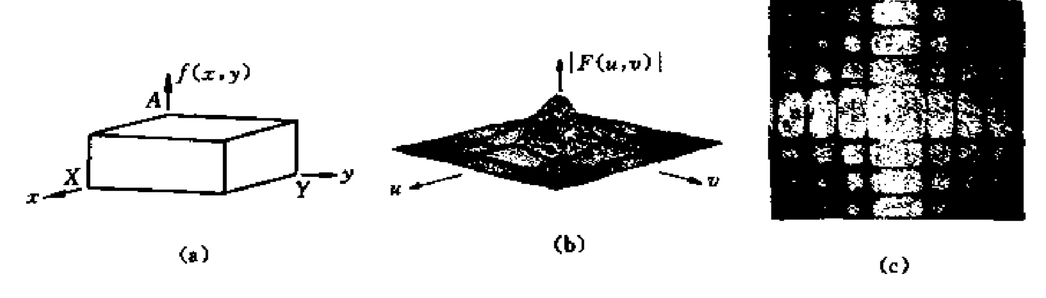



图 2-2 二维信号的傅里叶变换

(a)一个二维函数；(b)它的傅立叶谱；(c)作为强度函数显示的谱

图(b)是用三维透视法表示的 $F(u,v)$ 的图形。图(c)是作为强度函数表示的傅里叶谱，其亮度正比于 $|F(u,v)|$ 的幅度。当 $f(x,y)$ 图形的 $A$、$X$、$Y$ 等参数变化时其幅度谱也将相应地变化。当 $f(x,y)$ 在空间域旋转时，其频谱也将旋转。这些内容将在傅里叶变换的性质中详细叙述。

## §2.2 二维离散傅里叶变换

以上介绍的是连续信号的傅里叶变换，如果把连续信号加以抽样，变为离散信号，此离散信号的傅里叶变换已经在信号与系统课程中讲述，如果 $x(n)$ 为一数字序列，则其离散傅里叶正反变换由下式表示：

$$X(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\mathrm{e}^{-\mathrm{j}\frac{2\pi mn}{N}} \quad 其中\ m = 0,1,\cdots,N-1$$

$$x(n) = \sum_{m=0}^{N-1} X(m)\mathrm{e}^{\mathrm{j}\frac{2\pi mn}{N}} \quad 其中\ n = 0,1,\cdots,N-1 \tag{2.2.1}$$

其中，$N$ 为离散信号的周期，$n$ 为时间域变量，$m$ 为频率域变量，$2\pi/N$ 为基波频率 $\Omega_0$。以上正反变换都是在一个周期内取值，这是由于信号在离散化以后，无论原来是周期的，或是有限长序列的，它们的频谱都是离散的，并且以 $N$ 为周期，此 $N$ 可以是原周期序列的周期，或是有限长非周期序列周期开拓之后的周期，因此其时间域和频率域皆是离散的周期的。

对于二维离散傅里叶变换而言，其正反离散变换表示式为：

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\mathrm{e}^{[-\mathrm{j}2\pi(ux/M+vy/N)]}$$

式中，$u=0,1,2,\cdots,M-1$，$v=0,1,2,\cdots,N-1$

而

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\mathrm{e}^{[\mathrm{j}2\pi(ux/M+vy/N)]} \tag{2.2.2}$$

式中 $x=0,1,2,\cdots,M-1 \quad y=0,1,2,\cdots,N-1$

以上 $M$、$N$ 表示图像在 $x$、$y$ 方向具有不同大小的阵列，当图像阵列为方形时，则 $M=N$，则存在有

$$F(u,v) = \frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\mathrm{e}^{[-\mathrm{j}2\pi(ux/N+vy/N)]} \tag{2.2.3}$$

式中 $u=v=0,1,2,\cdots,N-1$

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)\mathrm{e}^{[+\mathrm{j}2\pi(ux/N+vy/N)]}$$

式中 $x=y=0,1,2,\cdots,N-1$

与连续的二维傅里叶变换一样，又称 $F(u,v)$ 为离散信号 $f(x,y)$ 频谱，$\varphi(u,v)$ 为相谱，$|F(u,v)|$ 为幅谱，其表示式为：

$$\begin{aligned} &F(u,v)=|F(u,v)|\mathrm{e}^{\mathrm{j}\varphi(u,v)}=R(u,v)+\mathrm{j}I(u,v) \\ &\varphi(u,v)=\mathrm{arctg}\,\frac{I(u,v)}{R(u,v)} \\ &|F(u,v)|=[R^2(u,v)+I^2(u,v)]^{1/2} \end{aligned} \tag{2.2.4}$$

由于实际上幅谱应用的较多，为此经常把幅谱称为频谱。

# § 2.3　二维离散傅里叶变换的性质

由于二维离散傅里叶变换的性质对图像的分析具有十分重要的作用，所以必须强调它的性质，而二维离散傅里叶变换又与二维连续傅里叶变换的性质基本上是对应的，所以二者对照地学习将十分有利于掌握。下面把其性质列出如下，如表 2－3－1 所示。

表 2－3－1　二维 DFT 的性质

| 编号 | 性质 | 表　示　式 |
|---|---|---|
| 1 | 线性性质 | $a_1f_1(x,y)+a_2f_2(x,y)\leftrightarrow a_1F_1(u,v)+a_2F_2(u,v)$ |
| 2 | 比例性质 | $f(ax,by)\leftrightarrow\frac{1}{\vert ab\vert}F(\frac{u}{a},\frac{v}{b})$ |
| 3 | 可分离性 | $F(u,v)=F_x\{F_y[f(x,y)]\}=F_y\{F_x[f(x,y)]\}$ |
|  |  | $f(x,y)=F_u^{-1}\{F_v^{-1}[F(u,v)]\}=F_v^{-1}\{F_u^{-1}[F(u,v)]\}$ |
| 4 | 空间位移 | $f(x-x_0,y-y_0)\leftrightarrow F(u,v)e^{-j2\pi(ux_0+vy_0)/N}$ |
| 5 | 频率位移 | $f(x,y)e^{j2\pi(u_0x+v_0y)/N}\leftrightarrow F(u-u_0,v-v_0)$ |
|  | 图像中心化 | 当 $u_0=v_0=\frac{N}{2}$ 时　$f(x,y)(-1)^{x+y}\leftrightarrow F(u-\frac{N}{2},v-\frac{N}{2})$ |
| 6 | 周期性 | $F(u,v)=F(u+aN,v+bN),f(x,y)=f(x+aN,y+bN)$ |
| 7 | 共轭对称性 | $f^*(x,y)\leftrightarrow F^*(-u,-v)$ |
| 8 | 旋转不变性 | $f(r,Q+Q_0)\leftrightarrow F(\rho,\varphi+Q_0)$ |
| 9 | 平均值 | $F(0,0)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)=\overline{f(x,y)}$ |
| 10 | 卷积定理 | $f(x,y)*h(x,y)\leftrightarrow F(u,v)H(u,v)$<br>$f(x,y)\cdot h(x,y)\leftrightarrow F(u,v)*H(u,v)$ |
| 11 | 相关定理 | 互相关　$f(x,y)\bigcirc g(x,y)\leftrightarrow F(u,v)G^*(u,v)$<br>$f(x,y)\cdot g^*(x,y)\leftrightarrow F(u,v)\bigcirc G(u,v)$ |
|  |  | 自相关　$f(x,y)\bigcirc f(x,y)\leftrightarrow\vert F(u,v)\vert^2$<br>$\vert f(x,y)\vert^2\leftrightarrow F(u,v)\bigcirc F(u,v)$ |
| 12 | 帕塞瓦定理<br>（能量定理） | $\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f_1(x,y)f_2^*(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F_1(u,v)F_2^*(u,v)$<br>当 $f_1(x,y)=f_2(x,y)$ 时<br>$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\vert f_1(x,y)\vert^2=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\vert F(u,v)\vert^2$ |

下面将以上性质中几个有重要意义的性质叙述一下。

1. 可分离性

它说明二维离散傅里叶变换可以分解为二个一维傅里叶变换来进行，而且其进行一维傅里叶变换可以先对 $x$ 后对 $y$，也可以先对 $y$ 后对 $x$，对顺序没有限制。同样对傅里叶反变换亦可以分成两步来求。

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1} e^{-j2\pi ux/N} \cdot \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi vy/N}$$

式中 $u,v=0,1,\cdots,N-1$，而

$$f(x,y) = \frac{1}{N}\sum_{u=0}^{N-1} e^{j2\pi ux/N} \cdot \sum_{v=0}^{N-1} F(u,v) e^{j2\pi vy/N} \tag{2.3.1}$$

式中 $x,y=0,1,2,\cdots,N-1$。

以上式子亦可以写为

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1} F(x,v) e^{-j2\pi ux/N}$$

其中

$$F(x,v)=N[\frac{1}{N}\sum_{y=0}^{N-1} f(x,y) e^{-j2\pi vy/N}] \tag{2.3.2}$$

$F(x,v)$说明对于 $f(x,y)$每一行取变换，将其结果乘以 $N$ 就得到二维函数 $F(x,v)$。然后再沿着 $F(x,v)$的每一列取变换就获得所需的结果 $F(u,v)$。同样，首先把 $f(x,y)$的诸列取变换，然后再将其结果的行取变换，将会得到同样的结果。

2. 频率位移性质

当图像在频率域希望移动时需用频率位移性质。

$$f(x,y) e^{j2\pi(u_0x+v_0y)/N} \leftrightarrow F(u-u_0,v-v_0) \tag{2.3.3}$$

上式说明 $f(x,y)$乘以一个指数项函数，并取其乘积的傅里叶变换，将使频率域的中心移动到$(u_0,v_0)$点，在把图像进行傅里叶变换之后，往往要把中心移动到 $u_0=v_0=N/2$ 的位置上，$N$ 为图像阵点的点数，在此情况下，可以得到：

$$e^{j2\pi(u_0x+v_0y)/N} = e^{j2\pi(x+y)\frac{N}{2}/N} = e^{j\pi(x+y)} = (-1)^{x+y}$$

则

$$f(x,y)(-1)^{x+y} \leftrightarrow F(u-\frac{N}{2},v-\frac{N}{2}) \tag{2.3.4}$$

这样可以在空间域把 $f(x,y)$乘以$(-1)^{x+y}$就将 $f(x,y)$傅里叶变换的原点移动到相应 $N\times N$ 频率方阵的中心，这个过程称之为图像中心化过程。以上过程如果在一维变量的情况下，它可简化为用$(-1)^x$ 项乘以 $f(x)$。以上过程是频率平移性质的应用，如果需要在时间域(或空间域)进行平移，同样可以运用空间位移性质，并要注意到，当图像在空间域平移时，其频域中的幅度谱并没有受到影响，而仅仅增加了相移项。此项内容希望同学们能够自己推导并理解。以下图 2-3 表示原空间域图像，其傅里叶变换幅度谱，以及坐标移动到$(N/2,N/2)$后求得的频谱。





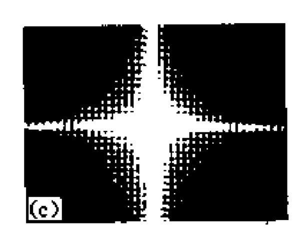



图 2-3 图像的中心化

(a)原空间域图像；(b)幅度谱；(c)中心化后的频谱

3. 周期性和共轭对称性

离散信号的傅里叶变换都有其周期性，如果以$(x+mN)$，$(y+nN)$代入反变换公式，以$(u+mN)$，$(v+nN)$代入正变换公式，不难对以上周期性加以证明，为此在时域、频域分析其特点，只需在其任何一个周期中的每个变量的 $N$ 个值就可以完全确定。

DFT 的共轭对称性亦是不难进行证明的：

$$F(u,v)=\frac{1}{M\cdot N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)e^{[-j2\pi(\frac{ux}{M}+\frac{vy}{N})]}$$

如果方程两边进行共轭

$$\begin{aligned}F^*(u,v)&=\frac{1}{M\cdot N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f^*(x,y)e^{[j2\pi(ux/M+vy/N)]}\\&=\frac{1}{M\cdot N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)e^{[-j2\pi(-ux/M-vy/N)]}\\&=F(-u,-v)\end{aligned}$$

以上推导过程中 $f(x,y)$必定是一个实函数

$$f^*(x,y)=f(x,y)$$

则存在 $F^*(u,v)=F(-u,-v)$，或 $f^*(x,y)\leftrightarrow F^*(-u,-v)$

二维 DFT 的周期性和共轭对称性给图形的频谱分析和显示带来了很大的好处，下面以一维的为例，周期性说明 $F(u)$具有长度为 $N$ 的周期，而对称性证明变换的幅度以原点为中心，如图 2-4 所示。

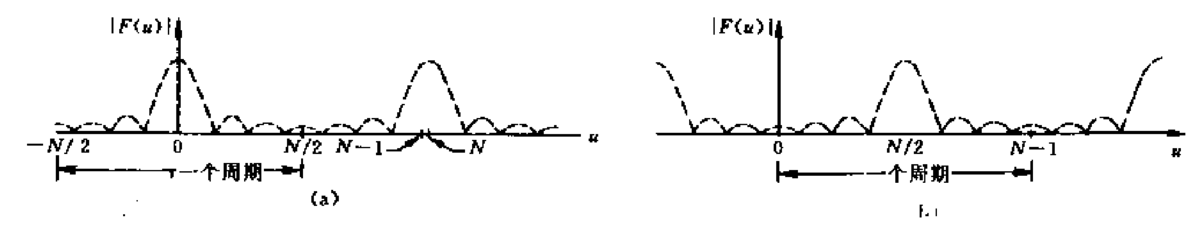



图 2-4 傅里叶变换的周期性

(a)表示在区间[0，$N-1$]中背对背半周期的傅里叶谱；(b)表示在同一区间内整个周期的平移之后的谱

$$F(u)=F(u+N),\qquad |F(u)|=|F(-u)|$$

由上图可以看出在$-N/2$ 到 $N/2-1$ 之间可以得到一个完整的频谱，而在 0 到 $N-1$ 之间只能得到该频谱相对的两半边，即前一周期的后一半和后一周期的前一半。由于计算 $DFT$ 的频率值已归一化为从 0 到 $N-1$，所以实际得到的是分离开的两半个频谱。为了在一个周期的变换中得到一个完整的频谱，可以把频率坐标的原点移动 $N/2$ 个采样间隔，即计算 $F(u+N/2)$。据定义

$$F(u+\frac{N}{2})=\frac{1}{N}\sum_{x=0}^{N-1}f(x)e^{-j\frac{2\pi}{N}x(u+\frac{N}{2})}=\frac{1}{N}\sum_{x=0}^{N-1}(-1)^x f(x)e^{-j\frac{2\pi}{N}xu}$$

为此，只要将被变换函数乘以$(-1)^x$ 再进行变换，便可在一个周期的变换中，求得一个完整的频谱。事实上，只须将被变换的数组，按顺序交替冠以正、负号，然后再取变换就可以了。

以上结论同样可以用于二维的情况，图 2-3 就是利用频域中心化性质，当 $f(x,y)$乘以$(-1)^{x+y}$以后，其频谱的中心原点将移动到$(N/2,N/2)$的位置，能量将从四角位置移动到中心位置，这也是以上周期性、共轭对称性的具体应用。

利用周期性和共轭对称性，在计算函数的频谱时，可以只对其一半的频谱进行分析与计算。

4. 旋转不变性

首先我们把 $x,y,u,v$ 皆用极坐标的形式表示

$$x=r\cos\theta \qquad y=r\sin\theta$$

$$u=\omega\cos\varphi \qquad v=\omega\sin\varphi$$

则 $f(x,y)$,$F(u,v)$可以分别由 $f(r,\theta)$和 $F(\omega,\varphi)$来表示,无论在连续的或离散的傅里叶变换中用直接代入法可以证明。

$$f(r,\theta+\theta_0)\leftrightarrow F(\omega,\varphi+\theta_0)$$

即当 $f(x,y)$被旋转 $\theta_0$,则 $F(u,v)$被旋转同一个角度。同样,如果 $F(\omega,\varphi)$被旋转一个角度,则 $f(r,\theta)$也被旋转同一个角度。

5. 平均值

二维离散函数普遍采用的平均值定义由下面的表达式给出

$$\bar{f}(x,y)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)$$

将 $u=v=0$ 代入公式(2.2.3)得

$$F(0,0)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)$$

因此 $\bar{f}(x,y)$与傅里叶变换系数的关系写为

$$\bar{f}(x,y)=F(0,0)$$

6. 离散卷积定理

设 $f(x,y)$、$g(x,y)$是大小分别为 $A\times B$ 和 $C\times D$ 的两个数组,则它们的离散卷积定义为

$$f(x,y)*g(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)g(x-m,y-n)$$

$$x=0,1\cdots M-1$$

$$y=0,1\cdots N-1$$

式中 $M=A+C-1$, $N=B+D-1$

现证明离散卷积定理,对上式两边取 DFT,则

$$\begin{aligned}F[f(x,y)*g(x,y)]&=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\Big\{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)g(x-m,y-n)\Big\}e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}\\&=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)e^{-j2\pi(\frac{um}{M}+\frac{vn}{N})}\times\\&\quad\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}g(x-m,y-n)e^{-j2\pi}\Big[\frac{u(x-m)}{M}+\frac{v(y-n)}{N}\Big]\\&=F(u,v)\cdot G(u,v)\end{aligned}$$

于是空间域的卷积定理得以证明。用类似方法也可以证明频率域之卷积定理。

7. 离散相关定理

大小为 $A\times B$ 和 $C\times D$ 的两个离散函数序列 $f(x,y)$,$g(x,y)$之互相关定义为

$$f(x,y)\bigcirc g(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)g(x+m,y+n)$$

式中:$M=A+C-1$

$N=B+D-1$

利用与卷积定理相似的方法，可以证明互相关和自相关定理。和连续的情况一样，它也说明了离散二维函数的自相关和自谱，两二维函数的互相关和互谱成傅里叶变换对的关系。

利用相关定理和二维 DFT 可以计算函数的相关，但和计算卷积一样，有循环相关的问题。为此，也必须将求相关的函数延拓成周期为 $M$ 和 $N$ 的周期函数，并对延拓后的函数添加适当的零点。

## §2.4 快速傅里叶变换(FFT)

对于有限长序列的离散傅里叶变换，有

$$X(m)=\sum_{n=0}^{N-1}x(n)\mathrm{e}^{-\mathrm{j}2\pi mn/N}\qquad(m=0,1,2,\cdots,N-1)$$

$$X(n)=\frac{1}{N}\sum_{m=0}^{N-1}X(m)\mathrm{e}^{\mathrm{j}2\pi mn/N}\qquad(n=0,1,2,\cdots,N-1)$$

令 $W=\mathrm{e}^{-\mathrm{j}2\pi/N}$ 和 $W^{-1}=\mathrm{e}^{\mathrm{j}2\pi/N}$ 可以简写为

$$X(m)=\sum_{n=0}^{N-1}x(n)W^{mn}\qquad x(n)=\frac{1}{N}\sum_{m=0}^{N-1}X(m)W^{-mn}$$

所谓快速傅里叶变换的算法亦就是在研究离散傅里叶变换计算的基础上，节省它的计算量，达到快速计算的目的，可以把以上变换的公式写成矩阵相乘的形式

$$\begin{bmatrix}X(0)\\X(1)\\\vdots\\X(N-1)\end{bmatrix}=\begin{bmatrix}W^{00} & W^{01} & \cdots & W^{0(N-1)}\\W^{10} & & \cdots & W^{0(N-1)}\\\vdots & & & \\W^{(N-1)0} & \cdots & \cdots & W^{(N-1)(N-1)}\end{bmatrix}\begin{bmatrix}x(0)\\x(1)\\\vdots\\x(N-1)\end{bmatrix}$$

在以上矩阵的计算中可以看出，每得到一个频率分量，必须进行 $N$ 次乘法和 $N-1$ 次加法的运算，要完成整个变换必须有 $N^2$ 次乘法和 $N(N-1)$次加法的运算，计算量比较大。

下面利用库利—图基计算方法设法减少乘法运算

设 $x_1(n)=x(2n)\qquad(n=0,1,\cdots,N/2-1$ 为偶数)

$x_2(n)=x(2n+1)\qquad(n=0,1,\cdots,N/2-1$ 为奇数)

为此傅里叶变换对可以写成如下的形式

$$X(m)=\sum_{n=0}^{N-1}x(n)W_N^{mn}=\sum_{n=0}^{\frac{N}{2}-1}x_1(n)W_N^{mn}+\sum_{n=0}^{\frac{N}{2}-1}x_2(n)W_N^{mn}$$

$$=\sum_{n=0}^{\frac{N}{2}-1}x(2n)W_N^{m(2n)}+\sum_{n=0}^{\frac{N}{2}-1}x(2n+1)W_N^{m(2n+1)}$$

其中第一项为偶部，第二项为奇部。因为

$$W_{2N}^{k}=W_N^{k/2}$$

所以

$$X(m)=\sum_{n=0}^{\frac{N}{2}-1}x(2n)W_{N/2}^{mn}+\sum_{n=0}^{\frac{N}{2}-1}x(2n+1)W_{N/2}^{mn}W_N^{m}$$

$$=\sum_{n=0}^{\frac{N}{2}-1}x(2n)W_{N/2}^{mn}+W_N^{m}\sum_{n=0}^{\frac{N}{2}-1}x(2n+1)W_{N/2}^{mn}=X_1(m)+W_N^{m}X_2(m)$$

式中 $X_1(m)$和 $X_2(m)$分别是 $x_1(n)$和 $x_2(n)$的 $N/2$ 点的傅里叶变换。$X_1(m)$和 $X_2(m)$是以 $N/$

2 为周期的，即

$$X_1(m+\frac{N}{2})=X_1(m) \qquad X_2(m+\frac{N}{2})=X_2(m)$$

这说明当 $m\geqslant N/2$ 时，上式也是重复的。

$$X(m)=X_1(m)+W_N^m X_2(m) \qquad (m=0,1,2,\cdots,N-1)$$

同样可以成立。可见一个 $N$ 点的离散傅里叶变换可以由二个 $N/2$ 点的傅里叶变换得到。这样分解之后乘法次数大大减少，总共需要 $2\times(N/2)^2+N$ 次乘法运算即可，原来需要 $N^2$ 次运算。可见分解以后乘法的计算次数减少了近一半。注意原来的计算区间是 $k\leqslant N-1$，而现在的计算区间是 $k\leqslant N/2-1$。当 $N$ 是 2 的整数幂时，则上式中的 $X_1(m)$ 和 $X_2(m)$ 还可以分成更短的序列，因此计算时间会更短。

下面以基数为 2，$N=8$ 的按时间分解的算法为例，请看图 2-5 的流程图。其输入是顺序的从 0 到 7，而输出是乱序的 0，4，2，6，1，5，3，7。

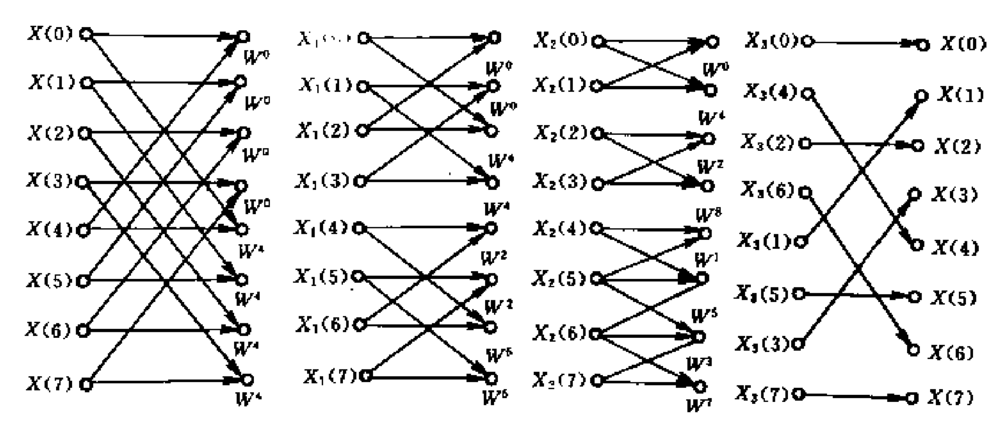


图 2-5 FFT 蝶式运算流程图(按时间分解)

先看：

$$\begin{aligned}X(1)&=x_2(4)+x_2(5)W^1\\&=x_1(4)+x_1(6)W^2+[x_1(5)+x_1(7)W^2]W^1\\&=x(0)+x(4)W^4+[x(2)+x(6)W^4]W^2+[x(1)+x(5)W^4]W^1+\\&\quad[x(3)+x(7)W^4]W^2W^1\\&=x(0)+x(4)W^4+x(2)W^2+x(6)W^4W^2+x(1)W^1+x(5)W^4W^1+\\&\quad x(3)W^2W^1+x(7)W^4W^2W^1\\&=x(0)+x(4)W^4+x(2)W^2+x(6)W^6+x(1)W^1+x(5)W^5+\\&\quad x(3)W^3+x(7)W^7\end{aligned}$$

而用计算公式所得的结果为

$$X(m)=\sum_{n=0}^{N-1}x(n)W_N^{mn}$$

其中 $X(1)=x(0)+x(1)W^1+x(2)W^2+x(3)W^3+x(4)W^4+x(5)W^5+x(6)W^6+x(7)W^7$

与前面用流程图计算的结果一致，以上是时间域的抽取法，将 $x(n)$ 进行分解，也还有将频域序列 $X(m)$ 进行分解的方法，这里不再详细谈了。

下面再谈一下用 FFT 进行傅里叶变换的过程：

用 FFT 进行傅里叶计算，需要解决以下四个问题，迭代次数 $r$ 的确定；对偶节点的计算，加权系数 $W_N^p$ 的计算以及重新排序问题。

从以上计算二维离散傅里叶变换公式

$$F(u,v)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)e^{[-j2\pi(ux+vy)/N]} \qquad (u,v=0\cdots\cdots N-1)$$

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}F(u,v)e^{[j2\pi(ux+vy)/N]} \qquad (x,y=0,1\cdots\cdots N-1)$$

(1)迭代次数 $r$ 的确定：$r=\log_2 N$ 其中 $N$ 为序列的长度，它是以 2 为基数的数，如果 $N=8$，则 $r=\log_2 8=3$，则需要 3 次迭代计算。

(2)对偶节点的计算：对偶节点的计算就是设法求出每次迭代中对偶节点的间隔或节距。在流程图中，把标有 $X_e(k)$ 的点称为节点，$k$ 为该点的序号，$e$ 为第 $e$ 次迭代。在图中可见第一次迭代的节距为 $N/2$，第二次迭代的节距为 $N/4$，第三次迭代的节距为 $N/2^3$ 等等，给出以下对偶节点的计算公式。

如果某节点为 $x_e(k)$，它的对偶节点为

$$x_e(k+\frac{N}{2^e})$$

其中 $e$ 是迭代的次数，$k$ 为序列的号数，$N$ 是序列的长度。

例：如果序列长度 $N=8$，求 $x_2(1)$ 的对偶节点

$$x_2(k+\frac{N}{2^e})=x_2(1+\frac{8}{2^2})=x_2(3)$$

从前面的流程图也可以验证这一点。

(3)加权系数 $W_N^p$ 的计算：在蝶形计算的过程中，有加权系数的计算，结合 $X_2(2)$ 加权系数 $W_8^p$ 实例，其步骤如下：

*a*. 把 $k$ 值写成 $r$ 位的二进制数，$k$ 是序列的序号数，$e$ 是迭代的次数，由 $x_2(2)$ 和 $W_8^p$ 可知 $e=2$，$k=2$，$N=8$，则 $r=\log_2 8=3$。故 $k=2$ 时写成三位的二进制数为 010。

*b*. 把此二进制数右移 $r-e$ 位，并把左边的空位补零，即把以上计算得的 010 向右移动 3－2位则结果为 001。

*c*. 把此右移的二进制数进行比特倒转，以上 001 则倒转为 100。

*d*. 把比特倒转后的二进制数翻成十进制数则为 $p$，以上 100 的十进制数为 4，则 $p=4$，结果 $x_2(2)$ 的加权值为 $W_8^4$

(4)重新排序：从前面的流程图可以看到，如果 $x(n)$ 是按顺序排列的，经过蝶形计算，其变换结果序列是非顺序的，即乱序的。反之，如果 $x(n)$ 是乱序的，则 $X(m)$ 是顺序的，为了更好地输出，需要将以上得到的变换序列再顺序过来，此重新整序的过程如下：

*a*. 将最后一次迭代的结果 $X_c(k)$ 中的序号数 $k$ 写成二进制数，即 $X_e(k)=X_r(k_{r-1},k_{r-2}\cdots k_1,k_0)$

*b*. 将 $r$ 位的二进制数比特倒置，即 $X_e(k_0,k_1,\cdots\cdots k_{r-2},k_{r-1})$。

*c*. 求出倒置后的二进制数代表的十进制数，就可以得到与 $X(k)$ 相对应 $X(m)$ 的序号数。

例：$X_3(0)\rightarrow X_3(000)\rightarrow X(000)\rightarrow X(0)$

$X_3(1)\rightarrow X_3(001)\rightarrow X(100)\rightarrow X(4)$

⋮　　　　⋮

$X_3(7)\rightarrow X_3(111)\rightarrow X(111)\rightarrow X(7)$

由以上蝶形计算可见，顺序可以变成倒序，倒序可以变成顺序。所以在迭代运算全部完成后，使用比特倒转法就可以得到正确的变换系数的顺序。

在编制整个计算程序时应当注意以下几点：

(1)由于 DFT 的计算都是复数运算，为此对于每个 $F(u,v)$的数都有实部和虚部两部分，为此计算过程中将求出其模(实部的平方与虚部平方和之开方)，其角(虚部与实部比之反正切)。最后以其幅模与频率量$(u,v)$的关系画出其幅谱图。

(2)由于幅谱是强度函数的显示，而图像的幅度强度函数衰减得很快，为此高频分量变得愈来愈不清楚，为解决此问题，常用 $D(u,v)$来代替 $F(u,v)$：

$$D(u,v)=\lg[1+|F(u,v)|]$$

注意：此处 $D(u,v)$为非负的函数，当$|F(u,v)|=0$ 时，$D(u,v)=0$。而 $F(u,v)$很高时，由于取对数之后，其高频与低频之差就变得很小，这样减少了高低频的幅度之差，使低频信息亦能够显示出来，便于人们的视觉理解。

(3)在具体实现二维离散 DFT 时，根据 DFT 的可分离性，可以对二维数组先进行列变换，然后再进行行变换。亦可以先对行进行变换，再进行列变换，然而这样做将给程序设计带来麻烦和浪费。为此可以在编程时，把第一次变换的结果进行转置，这是变动二维数组，而不是变动程序，仍按原程序进行变换，这样方便了运算。

(4)在进行二维傅里叶反变换时，只需将 $F(u,v)$共轭之后的 $F^*(u,v)$代入正变换的公式就得到了所求的 $f(x,y)$，请读者根据已学知识验证以上道理。注意在证明过程中，必须有 $f(x,y)$为实函数的假定，而这种假定往往是符合实际的。

有关二维 DFT 的计算实验，本章习题中有详细的要求以及具体细节，希望读者细细看，编制好程序，并请注意如何设计出最后的灰度图像，有关打印子程序请参考有关文献。

## §2.5 离散余弦变换(DCT)

一维离散余弦变换(DCT)的正向变换核由下式表示

$$g(x,0)=1/\sqrt{N}$$

$$g(x,u)=\sqrt{\frac{2}{N}}\cos\frac{(2x+1)u\pi}{2N}$$

其中 $x=0,1,\cdots,N-1$，以及 $u=1,2,\cdots,N-1$。将此结果代入通用的一维正变换的表示式

$$T(u)=\sum_{x=0}^{N-1}f(x)g(x,u)$$

其中 $T(u)$是 $f(x)$的变换，$g(x,u)$是正变换核，并且 $u$ 在 $0,1,2,\cdots,N-1$ 内取值，其反变换由下式表示：

$$f(x)=\sum_{u=0}^{N-1}T(u)h(x,u)$$

其中 $h(x,u)$是反变换核，并且 $x$ 的取值范围是 $0,1,2,\cdots,N-1$。则 $f(x)$的一维离散余弦变换 DCT 为

$$C(0)=\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}f(x)$$

$$C(u)=\sqrt{\frac{2}{N}}\sum_{x=0}^{N-1}f(x)\cos\frac{(2x+1)u\pi}{2N}$$

其中 $u=0,1,2,\cdots,N-1$，$c(u)$即是 $f(x)$的 DCT，其反向的 DCT 定义为

$$f(x)=\frac{1}{\sqrt{N}}C(0)+\sqrt{\frac{2}{N}}\sum_{u=1}^{N-1}C(u)\cos\frac{(2x+1)u\pi}{2N}$$

其中 $x=0,1,2,\cdots,N-1$。

把以上一维 DCT 扩展到二维离散变换，其变换核为

$$g(x,y,0,0)=\frac{1}{N}$$

$$g(x,y,u,v)=\frac{2}{N}\left[\cos\frac{(2x+1)\pi u}{2N}\right]\left[\cos\frac{(2y+1)\pi v}{2N}\right]$$

其中 $x,y=0,1,\cdots,N-1$ 而 $u,v=1,2,\cdots,N-1$。对于反变换核也是类似的形式。

对于二维 DCT 的通用形式

$$T(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)g(x,y,u,v)$$

以及

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}T(u,v)h(x,y,u,v)$$

其中 $g(x,y,u,v)$和 $h(x,y,u,v)$为二维 DCT 的正反变换核。把以上变换核式代入其中，则得

$$C(0,0)=\frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)$$

$$C(u,v)=\frac{2}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\left[\cos\frac{(2x+1)\pi u}{2N}\right]\left[\cos\frac{(2y+1)\pi v}{2N}\right]$$

其中 $u,v=0,1,2,\cdots,N-1$。其反变换为

$$f(x,y)=\frac{1}{N}C(0,0)+\frac{2}{N}\sum_{u=1}^{N-1}\sum_{v=1}^{N-1}C(u,v)\left[\cos\frac{(2x+1)\pi u}{2N}\right]\left[\cos\frac{(2y+1)\pi v}{2N}\right]$$

其中 $x,y=0,1,\cdots,N-1$。

根据变换核的可分离性，二维的正向或反向变换能够逐次应用一维 DCT 的算法加以计算。

有关 DCT 的实际算法，有一个十分有趣而简便的算法，它可以直接由 FFT 的算法来求得。

将以上的离散余弦变换式可以变成以下的形式

$$C(0)=\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}f(x)$$

$$C(u)=\sqrt{\frac{2}{N}}\mathrm{Re}\left\{\left[\exp\left(\frac{-\mathrm{j}2\pi u}{2N}\right)\right]\times\sum_{x=0}^{2N-1}f(x)\exp\left(\frac{-\mathrm{j}2\pi ux}{2N}\right)\right\}\qquad(u=1,2,\cdots,N-1)$$

其中，对 $x=N,N+1,\cdots,2N-1$ 点处有 $f(x)=0$，即将原来 $N$ 个点扩展为 $2N$ 个点，补之以零值点。公式中求和项就是 $2N$ 个点的离散傅里叶变换，Re 即代表括号内的项的实部。为此可以用 FFT 程序计算出 DCT。类似地用 $2N$ 个点上的反 FFT 也可以从 $c(u)$中求出 $f(x)$。

## §2.6　离散沃尔什变换和哈达玛变换

离散傅里叶变换和余弦变换在快速算法中要用到复数乘法，占用的时间仍然比较多。在某

些应用领域中，需要更为便利更为有效的变换方法。沃尔什变换就是其中的一种。沃尔什函数是在 1923 年由美国数学家沃尔什提出来的。在沃尔什的原始论文中，给出了沃尔什函数的递推公式，这个公式是按照函数的序数在正交区间内过零点的平均数来定义的。不久以后，这种规定函数序数的方法也被波兰数学家卡兹马兹采用了，所以通常将这种规定函数序数的方法称为沃尔什-卡兹马兹定序法。

1931 年美国数学家佩利又给沃尔什函数提出了一个新的定义。他指出，沃尔什函数可以用有限个拉格尔函数的乘积来表示，这样得到的函数的序数与沃尔什得到的函数的序数是完全不同的。这种定序方法是用二进制来定序的，所以称为二进制序数或自然序数。

利用包含 +1 和 −1 阵元的正交矩阵可以将沃尔什函数表示为矩阵的形式。早在 1867 年，英国数学家希尔威斯特已经研究过这种矩阵。后来法国数学家哈达玛在 1893 年将这种矩阵加以普遍化，建立了所谓哈达玛矩阵。利用克罗内克乘积算子不难把沃尔什函数表示为哈达玛矩阵的形式。利用这种形式定义的沃尔什函数称为克罗内克序数。这就是沃尔什函数的第三种定序法。

由上述历史可见，沃尔什函数及其有关函数的数学基础早就奠定了。但是，这些函数在工程中得到应用却是近几十年的事情。主要在于半导体器件和计算机的迅速发展，它的发展为沃尔什函数的实用解决了手段问题，因此也使沃尔什函数得到了进一步的发展。与傅里叶变换相比，沃尔什变换的主要优点在于减少存贮空间和提高运算速度。这一点对图像处理来说是至关重要的。特别是在大量数据需要进行实时处理时，沃尔什函数就更加显示出它的优越性。

1. *正交函数的概念*

一组实值的连续函数 $\{S_n(t)\}=\{s_0(t),s_1(t),s_2(t)\}$ 在 $0\leqslant t\leqslant T$ 区间内，如满足下式

$$\int_0^T kS_n(t)\cdot S_m(t)\mathrm{d}t=\begin{cases}Z & m=n\\ 0 & m\neq n\end{cases} \tag{2.6.1}$$

则 $\{s_n(t)\}$ 在区间 $0\leqslant t\leqslant T$ 内是正交的。$m,n$ 是正整数，$k$ 是与 $m,n$ 无关的非负的常数。如果 $k=1$，则称为归一化正交。任一组非归一化的正交函数总可以变换为归一化的正交函数，但如果 $f(t)$ 是定义在 $(0,T)$ 区间上的实值信号，利用正交函数可以表示为下式

$$f(t)=\sum_{n=0}^{\infty}C_nS_n(t) \tag{2.6.2}$$

式中，$C_n$ 是第 $n$ 项的系数。

一组完备的正交函数必然是一组闭合函数。完备的正交函数必须满足下述两个条件。第一是不存在这样一函数 $f(t)$，它满足

$$\int_T f^2(t)\mathrm{d}t<\infty \tag{2.6.3}$$

而且也满足 $\int_T f(t)S_n(t)\mathrm{d}t=0,\quad n=0,1,2,\cdots$。这个条件的意思是说，再也没有不属于 $\{S_n(t)\}$ 的某个非零的函数 $f(t)$，它与 $\{S_n(t)\}$ 的每一个函数正交。也就是说，所有互相正交的函数都包括在 $\{S_n(t)\}$ 的里面了。

第二是对于任何满足 $\int_T f^2(t)\mathrm{d}t<\infty$ 的函数，对给定的任意微小正数 $\epsilon=0$，总存在一个正整数 $N$ 与有限展开式

$$f_1(t)=\sum_{n=0}^{N-1}C_nS_n(t) \tag{2.6.4}$$

使得
$$\int_T |f(t)-f_1(t)|^2 dt < \varepsilon$$
这一条件的意思是，在一个能量有限的信号内总可以用有限级数来逼近它。对于给定的误差 $\varepsilon$ 来说，总可以找到一个 $N$ 值，使这种逼近的精确度满足要求。

完备性的必要和充分条件是在正交区间内各分量函数的平方之和存在，并且应当完全满足帕塞瓦尔定理。这个条件的物理意义是一组完备的正交函数所包含的能量，无论是时域，还是在变换域中都是相同的。完备性的重要意义在于：只有当正交函数是完备的，我们才能将一个满足一定条件的函数展开为此正交函数集的级数，否则将不能保证能量相等。

正交性、完备性、归一化的定义适用于一组函数中的所有函数，它们规定的区间可以是半无限区间 $(0,\infty)$，也可以是全无限区间 $(-\infty,\infty)$，当然也可以是有限区间 $(-T/2,T/2)$，$(0,T)$ 等等。

沃尔什函数的一个有用的特点是由有限个沃尔什函数组成，时间受限信号在变换域中只占据有限区间。而对于一般函数来说则不然，一般来说，时间受限，其频谱区间是无限的。反之，频域受限信号，在时间域上则是无限的。

2. 拉格尔函数

拉格尔函数集是一个不完备的正交函数集。由它可以构成完备的沃尔什函数。拉格尔函数包括 $n$ 和 $t$ 两个自变量，用 $R(n,t)$ 来表示。把一个正弦函数作无限限幅就可以得到拉格尔函数。它可用下式来表示：

$$R(n,t)=\mathrm{sgn}(\sin 2^n \pi t)$$
$$\mathrm{sgn}(x)=\begin{cases}1 & x>0\\ -1 & x<0\end{cases} \tag{2.6.5}$$

当 $x=0$ 时，$\mathrm{sgn}(x)$ 无定义。

由 sin 函数的周期性知道 $R(n,t)$ 也是周期性函数。由上式可见，当 $n=1$ 时，$R(n,t)$ 的周期为 1，$n=2$ 时，$R(2,t)$ 的周期为 1/2；当 $n=3$ 时，$R(3,t)$ 的周期为 $1/2^2$；一般情况下可用下式表示：

$$R(n,t)=R(n,t+\frac{1}{2^{n-1}}) \qquad (n=1,2,\cdots\cdots)$$

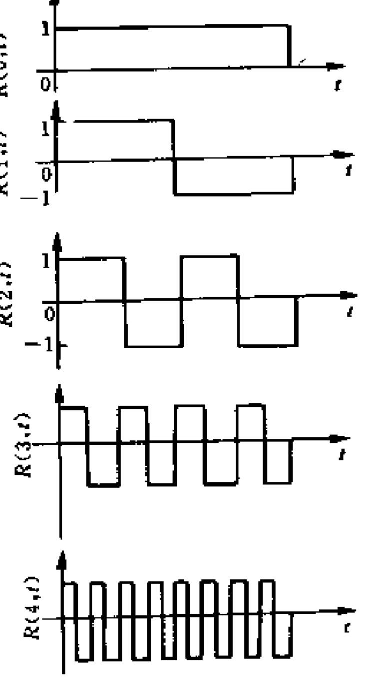


图 2－6　拉格尔函数

拉格尔函数的波形如图 2－6 示，由图 2－6 可见，拉格尔函数有如下一些规律：

(1) $R(n,t)$ 的取值只有 +1 和 −1。

(2) $R(n,t)$ 是 $R(n-1,t)$ 的二倍频。因此最高次数 $m=n$，则其它拉格尔函数可用脉冲分频器来产生。

(3) 如果已知 $n$，那么 $R(n,t)$ 有 $2^{n-1}$ 个周期，其中 $0<t<1$，如果在 $t=k+\frac{1}{2}\left(\frac{1}{2^n}\right)$ 处作取样，则可得到一数据序列 $R(n,k)$。

$k=0,1,2\cdots2^{n-1}$ 每一取样序列将与下述矩阵相对应。这里取 $n=3$，$k=0,1,2,\cdots,7$，得

$$\begin{bmatrix}R(0,k)\\R(1,k)\\R(2,k)\\R(3,k)\end{bmatrix}=\begin{bmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&-1&-1&-1&-1\\1&1&-1&-1&1&1&-1&-1\\1&-1&1&-1&1&-1&1&-1\end{bmatrix} \tag{2.6.6}$$

采用上述离散矩阵形式就可以用计算机进行灵活处理。

3. 沃尔什函数

沃尔什函数是完备的正交函数集，它的值也只取+1 和−1。从排队次序来定义不外有三种：一种是按沃尔什排列或按列率排队来定义；第二种是按佩利排队或称自然排列来定义；第三种是按哈达玛排列来定义。还可用其它方式来定义，但至今尚未统一。下面仅讨论第一种排列方法定义的沃尔什函数。

按沃尔什排列的沃尔什函数用 $wal(i,t)$ 表示，函数波形如图 2-7 所示。此沃尔什函数实际上就是按列率排列的。通常把正交区间内的波形变号次数的 $\frac{1}{2}$ 称为列率。如果令 $i$ 为波形在正交区间内的变号次数。那么按照 $i$ 为奇数或偶数。函数 $wal(i,t)$ 的列率将分别由下式决定。

$$Si=\begin{cases}0 & i=0\\ \frac{i+1}{2} & i=\text{奇数}\\ \frac{i}{2} & i=\text{偶数}\end{cases} \qquad (2.6.7)$$

按沃尔什排列的沃尔什函数可由拉格尔函数构成，它的表达式如下：

$$wal(i,t)=\prod_{k=0}^{p+1}[R(k+1,t)]^{g(i)_k} \qquad (2.6.8)$$

式中 $R(k+1,t)$ 是拉格尔函数，$g(i)$ 为 $i$ 的格雷码，$g(i)_k$ 是此格雷码的第 $k$ 位数字，$p$ 为正整数。

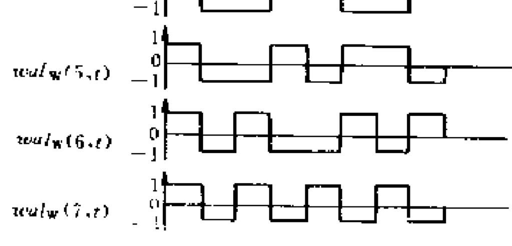


图 2-7 按沃尔什排列的沃尔什函数

一个正整数可以编成自然二进码，也可以编成格雷码。格雷码的特点是：两个相邻数的格雷码只有一个码位数值不同。例如 2 的格雷码是(0011)，3 的格雷码是(0010)。这两个相邻的数字的格雷码只有第四个码位的值不同。在脉冲编码技术中，常常采用这种码，以便得到较好的误差特性。一个正整数的自然二进码和格雷码之间是可以互相转换的。从自然二进码转成格雷码的方法如下：设一个十进制数 $n$ 的自然二进码为 $(n_{p-1},n_{p-2},\cdots,n_2,n_1,n_0)$，设 $n$ 的格雷码为 $(g_{p-1},g_{p-2},\cdots,g_2,g_1,g_0)$。其中 $n_k$ 和 $g_k$ 分别为二进码和格雷码内的码位数字，并且 $n_k,g_k\in\{0,1\}$。它们之间的关系可用下式表示

$$\begin{aligned}g_{p-1}&=n_{p-1}\\ g_{p-2}&=n_{p-1}\oplus n_{p-2}\\ g_{p-3}&=n_{p-2}\oplus n_{p-3}\\ g_k&=n_{k+1}\oplus n_k\\ g_1&=n_2\oplus n_1\\ g_0&=n_1\oplus n_0\end{aligned}$$

式中 $\oplus$ 代表模 2 加。

例如：$n(2)=(0010)$ 二进，其中 $n_3=0,n_2=0,n_1=1,n_0=0$，所以

$$g_3=n_3=0$$

$$g_2=n_3\oplus n_2=0\oplus 0=0$$

$$g_1=n_2\oplus n_1=0\oplus 1=1$$
$$g_0=n_1\oplus n_0=1\oplus 0=1$$

其格雷码为 $g(2)=(0011)_{格雷}$

同理，若 $n(3)=(0011)_{二进}$，则其格雷码为

$$g(3)=(0010)_{格雷}$$

在格雷码中，有如下关系存在

$$g(m)+g(n)=g(m\oplus n)$$

下面再根据式(2.6.8)写出沃尔什函数。现求 $p=4$ 时的 $wal(5,t)$，由于 $i=5$，所以 5 的二进码为(0101)，由前面叙述的转换规律，可知道其格雷码为(0111)。

| $(0$ | $1$ | $1$ | $1)_{格雷}$ |
|---|---|---|---|
| 第 3 位 | 第 2 位 | 第 1 位 | 第 0 位 |
| $g(5)_3$ | $g(5)_2$ | $g(5)_1$ | $g(5)_0$ |

即 $g(5)_0=1$， $g(5)_1=1$， $g(5)_2=1$， $g(5)_3=0$

将以上结果代入式(2.6.8)得

$$wal(5,t)=[R(1,t)]^1[R(2,t)]^1[R(3,t)]^1[R(4,t)]^0$$
$$=R(1,t)\cdot R(2,t)\cdot R(3,t)$$

例：当 $p=4,i=9$ 时，求 $wal(9,t)$，即

$$g(9)_3=1,\quad g(9)_2=1,\quad g(9)_1=0,\quad g(9)_0=1$$

代入式(2.6.8)得

$$wal(9,t)=[R(1,t)]^1\cdot[R(2,t)]^0\cdot[R(3,t)]^1\cdot[R(4,t)]^1$$
$$=R(1,t)\cdot R(3,t)\cdot R(4,t)$$

4. 沃尔什变换

离散沃尔什变换可由下列二式表达

$$w(i)=\frac{1}{N}\sum_{t=0}^{N-1}f(t)\cdot wal(i,t)$$
$$f(t)=\sum_{i=0}^{N-1}w(i)\cdot wal(i,t)$$

离散沃尔什变换亦可以写成矩阵的形式

$$\begin{bmatrix}w(0)\\ \vdots\\ w(N-1)\end{bmatrix}=\frac{1}{N}[wal(N)]\begin{bmatrix}f(0)\\ \vdots\\ f(N-1)\end{bmatrix}$$

$$\begin{bmatrix}f(0)\\ f(1)\\ \vdots\\ f(N-1)\end{bmatrix}=[wal(N)]\begin{bmatrix}w(0)\\ w(1)\\ \vdots\\ w(N-1)\end{bmatrix}$$

式中 $wal(N)$ 代表 $N$ 阶沃尔什矩阵。

5. 离散的哈达玛变换

哈达玛变换与沃尔什变换相比，没有本质上的不同，只是排列顺序的不同。但由于哈达玛矩阵具有最简单的递推关系，也就是高阶矩阵可以用低阶矩阵直接得到，这一特点使人们更愿

意采用哈达玛变换。

哈达玛变换是由哈达玛矩阵引出的。这种矩阵是用+1和−1作为阵元的$(N\times N)$阶方阵，它的行与列是正交的。

最小尺寸的正交哈达玛矩阵是2×2的，可表示为

$$H_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}h_{00} & h_{01}\\ h_{10} & h_{11}\end{bmatrix}=\frac{1}{\sqrt{2}}\{h_{ij}\}\qquad i,j=0,1 \tag{2.6.11}$$

矩阵中的阵元

$$h_{ij}=(-1)^{i,j} \tag{2.6.12}$$

目前，$N$数可达到200，$N=2^n$，$n$是整数。此时尺寸为$2N$的$H$矩阵为

$$H_{2N}=\frac{1}{\sqrt{2}}\left[\begin{array}{c|c}H_N & H_N\\ \hline H_N & \overline{H_N}\end{array}\right]=H_N\oplus H_N; \tag{2.6.13}$$

例如

$$H_4=H_2\oplus H_2=\frac{1}{\sqrt{4}}\left[\begin{array}{cc|cc}1 & 1 & 1 & 1\\ 1 & -1 & 1 & -1\\ 1 & 1 & -1 & -1\\ 1 & -1 & -1 & 1\end{array}\right] \tag{2.6.14}$$

列内符号变化的次数分别为0、3、1、2。这里矩阵中的阵元$h_{ij}$，可用下式表达

$$h_{ij}=(-1)^{i,j}=(-1)^{i_0,j_0\oplus i_1,j_1} \tag{2.6.15}$$

式中的行$i(i=0,1,2,3)$列$j(j=0,1,2,3)$用二进制数码表达为

$$i=i_1,i_0,\quad j=j_1,j_0$$

其中$h_{22}$，由于$i=10$，$j=10$，故$i_1=j_1=1$，$i_0=j_0=0$，因此得

$$h_{22}=(-1)^{(0,0)\oplus(1,1)}=-1$$

其中$h_{21}$，由于$i=10$，$j=01$，故$i_1=1$，$i_0=0$，$j_1=0$，$j_0=1$，因此得

$$h_{21}=(-1)^{(0,1)\oplus(1,0)}=1$$

推而广之，当$N=2^n$时，行$i$和列$j$可表达为

$$i=i_{n-1}i_{n-2}\cdots i_2i_1i_0$$

$$j=j_{n-1}j_{n-2}\cdots j_2j_1j_0$$

对应的$H_N$中的矩阵阵元为

$$h_{ij}=(-1)^P$$

而

$$P=\sum_{k=0}^{p-1}i_kj_k \tag{2.6.16}$$

上式中的加法是模2加⊕。由此可以看出，哈达玛矩阵中的阵元，在计算使用中不必预先存贮起来，它可以由计算机很容易地生成出来。

哈达玛矩阵中的+1、−1可以看作是从某种连续矩形波中采样出来的样本，其采样周期为$1/N$。这种连续矩形波函数就是沃尔什函数。哈达玛变换就是用一组矩形波来实施函数的分解，而傅里叶变换则用一组正弦波来实施函数的分解。设令$N=2^n$，则使用$(N^2\times N^2)$的对称$H$矩阵作变换时，可得矢量矩阵变换关系式为

$$g=Hf\qquad H=H_N\oplus H_N \tag{2.6.17}$$

由于$H$矩阵是正交，可分离的，而且是对称的，因此可得哈达玛变换的矩阵表达式为

$$G=H_NFH_N \tag{2.6.18}$$

而其逆变换由于 $H_N$ 的正交及对称性质，可得

$$F=H_NGH_N \tag{2.6.19}$$

哈达玛变换可以写成级数的表达形式如下：

$$G(u,v)=\frac{1}{N}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}F(j,k)\ (-1)^{p(j,k,u,v)} \tag{2.6.20}$$

式中 $p(j,k,u,v)=\sum_{i=0}^{n-1}(u_ij_i+v_ik_i)$。这里的 $u_i,v_i,j_i,k_i$ 是 $u,v,j,k$ 值的二进制数码表达式中第 $i$ 位的状态值，求和式的操作是模 2 加。

采用哈达玛变换，可以在计算中只作变号及加减法运算，而避开了傅里叶变换中使用的那种复数乘法运算。这种变换的运算也有快速算法，下面举例说明。

例：若 $f$ 为(4×1)的矢量，则变换式为

$$F=H_4f \tag{2.6.21}$$

展开后可得

$$\begin{bmatrix}F_0\\F_1\\F_2\\F_3\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&-1&1&-1\\1&1&-1&-1\\1&-1&-1&1\end{bmatrix}\begin{bmatrix}f_0\\f_1\\f_2\\f_3\end{bmatrix} \tag{2.6.22}$$

直接计算这个变换式需要 $N(N-1)$ 次加减运算，即

$$F_0=\frac{1}{2}[f_0+f_1+f_2+f_3]$$

$$F_1=\frac{1}{2}[f_0-f_1+f_2-f_3]$$

$$F_2=\frac{1}{2}[f_0+f_1-f_2-f_3]$$

$$F_3=\frac{1}{2}[f_0-f_1-f_2+f_3]$$

如果序贯分步地作变换计算，可以节省运算次数，如先作 $a_0=f_0+f_2,a_1=f_0-f_2,a_2=f_1+f_3$，$a_3=f_1-f_3$，然后作下一步运算可得

$$F_0=\frac{1}{2}[a_0+a_2],F_1=\frac{1}{2}[a_0-a_2],$$

$$F_2=\frac{1}{2}[a_1+a_3],F_3=\frac{1}{2}[a_1-a_3] \tag{2.6.24}$$

这种序贯分布的变换运算只需要 $N\log_2N$ 次即 8 次加减法运算，显然这样可以节省运算量。对于 $N=2^n$，它可以分成几步运算，变换运算量只需要 $N\log_2N$ 次即可。因此哈达玛运算具有与快速傅里叶变换类似的快速算法。图 2－8 示出 $N=4$ 的快速算法流程图。

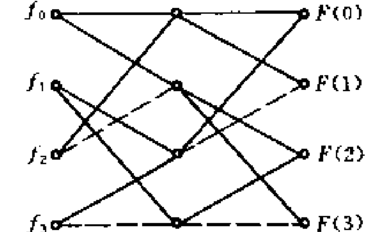


图 2－8　快速算法流程图

## §2.7　子波变换

子波变换的理论是近年来发展起来的，它目前已成为国际上极为活跃的研究领域。

长久以来，傅里叶分析是信号数据处理的有力工具，特别是在频谱分析和滤波方法的分析上无处不用到傅里叶分析。人们一般用函数来描述信号，总是把时间或空间作为自变量，而把

反映信号的物理量作为函数。但是傅里叶变换反映的是信号或函数的整体特征，而有些实际问题中需要关心的却是信号在局部范围中的特征。例如，在音乐和语言信号中人们关心的是什么时刻演奏什么音符，发出什么样的音节；对地震记录来说，人们关心的是什么位置出现什么样的反射波；在图形识别中的边缘检测，关心的是信号突变部分的位置。为了弥补傅里叶变换这方面的不足，1946 年 Gabor 引进了窗口傅里叶变换的概念。他用一个有限区间(称为窗口)外恒等于零的光滑函数(这个有限区间的位置随一个参数而变)去乘所要研究的函数，然后对它作傅里叶变换。这种变换确实能反映函数在窗口内部分的频率特性，因而确能在一些需要研究信号的局部性质的问题中起一定的作用。但是 Gabor 引入的这种变换窗口的窗口尺寸和形状却与频率无关而是固定不变的。这与高频信号的分辨率应比低频信号高，因而频率愈高窗口应愈小这一要求不符，为此它未能得到广泛的应用与发展。

子波变换继承和发展了 Gabor 的窗口傅里叶变换局部化思想，但它的窗口随频率增高而缩小，符合了高频信号的分辨力较高的要求，为此得到了迅速的发展，其理论逐步完善，应用面也愈来愈广。

## 一、加窗傅里叶变换

前已叙述，经典傅里叶变换是以全时域作为整体进行分析的，信号的时变特性得不到体现，为了分析非平衡信号，或信号的局域特性，人们发展了信号的时频分析法，加窗傅里叶变换(也称短时傅氏变换)就是其中的一种，它把非平稳信号看成是一系列短时平稳信号的叠加，而短时性则是通过时域上的加窗来获得，$f(x)\in L^2(R)$的加窗傅氏变换定义为

$$F(\omega,t)=\int_{-\infty}^{\infty}e^{-j\omega x}g(x-t)f(x)dx \tag{2.7.1}$$

其中 $g(x)$称为窗函数。一般情况下，$g(x)$为实函数且其傅氏变换的能量集中在低频处，它还可以看作是一低通滤波的脉冲响应，为了归一化，一般取

$$\|g\|^2=\int_{-\infty}^{\infty}|g(x)|^2dx=1 \tag{2.7.2}$$

式(2.7.1)定义的加窗傅里叶变换具有如下性质：

1)它从 $L^2(R)$到 $L^2(R^2)$是等距的，即

$$\int_{-\infty}^{\infty}|f(x)|^2dx=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(\omega,t)|^2d\omega dt \tag{2.7.3}$$

2)可以由 $F(\omega,t)$重构 $f(x)$，即

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\omega,t)g(t-x)e^{j\omega x}d\omega dt \tag{2.7.4}$$

它可以由 parseval 定理与式(2.7.1)证明之。

3)加窗傅里叶变换是一信号的冗余表示，而且是稳定完备的。

4)如果对所有$(\omega,t)\in R^2$ 均匀采样，离散加窗傅里叶变换可定义为$\forall\ (n,m)\in Z^2$，且

$$\begin{aligned}DF(m,n)&=DF(m\omega_0,nt_0)\\&=\int_{-\infty}^{\infty}e^{jm\omega_0 x}g(x-nt_0)f(x)dx\end{aligned} \tag{2.7.5}$$

采样间隔 $t_0$ 和 $\omega_0$ 的选择必须覆盖整个相空间，由采样集合$\{DF(m,n)\}(n,m)\in Z^2$，重构$f(x)\in L^2(R)$ 的条件是算子 $D:L^2(R)\rightarrow l^2(Z^2)$有界可逆，且有

$$\mathrm{DF}(m,n)=\langle f(x),e^{jm\omega_0 x}g(x-nt_0)\rangle \tag{2.7.6}$$

5)加窗傅里叶分析的空域和频率分辨率是常数，在空域这一分解表示所提供的信息在 $\sigma_t$（为窗函数 $g(x)$ 的标准偏差）内是非局域化的。

加窗傅里叶变换是一种局域化的时频分析方法，只要适当地选择窗函数 $g(x)$，就可以通过信号 $f(x)$ 的加窗傅里叶变换 $F(\omega,t)$ 获得它在 $2t_0$ 时间区间内的信息。这表明：一方面可以用加窗傅里叶变换来分析信号 $f(x)$ 的局部性质；另一方面，一旦窗函数 $g(x)$ 取定，其窗口大小也随之取定，因此通过窗口傅里叶变换只能获得 $f(x)$ 在窗口时间区间 $2t_0$ 内的信息。这样对短时高频信号，用加窗傅里叶变换难以获得希望的结果。固然可以通过缩小时窗宽度和采样步长 $t_0$ 改进加窗傅里叶分析，但窗口太窄会降低频率分辨率，对低频分量也不合适，同时会使计算变得相当复杂，这就是关于时频关系的不确定性原理，时间分辨率和频率分辨率不可能同时达到最佳。由此可知，如欲提高频率分辨率，窗口的大小应随频率而变，频率愈高，窗口应愈小。

## 二、连续子波变换

子波分析的基本思想是用一族函数去表示或逼近一信号或函数，这族函数称为子波函数集，它是通过一基本子波函数的不同尺度的平移和伸缩构成。子波函数集的特点是它的时宽频宽乘积很小，且在时间和频率轴上都很集中。

若基本子波函数为 $h(x)$，伸缩和平移因子分别为 $a$ 和 $b$，则子波变换基底定义为

$$h_{a,b}(x)=|a|^{1/2}h\left(\frac{x-b}{a}\right) \tag{2.7.7}$$

函数 $f(x)\in L^2(R)$ 的连续子波变换定义为

$$w_{a,b}=\int_{-\infty}^{\infty}h_{a,b}(x)f(x)\mathrm{d}x \tag{2.7.8}$$

写成内积形式即有

$$w_{a,b}=\langle f(x),h_{a,b}(x)\rangle \tag{2.7.9}$$

它对应于 $f(x)\in L^2(R)$ 在函数族 $h_{a,b}(x)$ 上的分解，这一分解必须满足下列可容许性条件

$$w_h=\int_0^{\infty}\frac{|H(\omega)|^2}{\omega}\mathrm{d}\omega<\infty \qquad 或 \qquad \int_{-\infty}^{\infty}h(t)\mathrm{d}t=0 \tag{2.7.10}$$

这里 $H(\omega)$ 是 $h(x)$ 的傅里叶变换，$h(x)$ 为带通滤波器的脉冲响应，因此子波变换式(2.7.8)可描述为函数 $f(x)\in L^2(R)$ 通过一带通滤波器的滤波。由 $w_{a,b}$ 重构 $f(x)$ 的子波逆变换定义为

$$f(x)=\frac{1}{w_h}\int_{-\infty}^{\infty}\int_0^{\infty}w_{a,b}h_{a,b}(x)\mathrm{d}a\mathrm{d}b \tag{2.7.11}$$

设 $\omega_0$ 是 $H(\omega)$ 的通带中心频率，即 $\int_0^{\infty}(\omega-\omega_0)|H(\omega)|^2d\omega=0$，而 $\sigma_\omega$ 是关于 $\omega_0$ 的均方根值带宽，即

$$\sigma_\omega^2=\int_0^{\infty}(\omega-\omega_0)^2|H(\omega)|^2\mathrm{d}\omega \tag{2.7.12}$$

很显然，$H_{a,b}(\omega)$ 的通带中心是 $a\omega_0$，带宽是 $a\sigma_\omega$。在对数坐标中，$H_{a,b}(\omega)$ 的带宽对所有 $a\in R$ 是一致的。所以子波变换将一信号分解为对数坐标中一具有相同大小的多通道频带集合。换句话说，子波 $h_{a,b}(x)$ 在频域中的能量集中于 $a\omega_0$，标准方差为 $a\sigma_\omega$，与加窗傅里叶变换不同，子波分辨率随标尺因子 $a$ 的变化而变化，当 $a$ 较小时，频域分辨性能较差，而时域分辨较好；当 $a$ 增加时，频域分辨率增加，而对时域分辨率则减小。子波分辨率的这种变化规律使得它具有良好

的局域性和非正则过零特性，因此它用于突变信号和奇异信号与图像的检测。

应当指出，$h_{a,b}(x)$不是正交的，其冗余性很大，此外子波基并不是唯一的，只要满足可容性条件即可定义特定信号的子波基。由式(2.7.8)定义的连续子波变换具有如下性质：

1)子波变换是一线性运算，它把一信号分解成不同尺度的分量。

2)子波是一函数 $h(x)\in L^2$，且$\int_{-\infty}^{\infty}h(x)\mathrm{d}x=0$，进一步对函数 $f(x)\in L^2(R)$，其子波变换可表示为

$$W_{a,b}=f*h_{a,b} \tag{2.7.13}$$

它可以看作是求函数 $f(x)$在 $h(x)$的各种尺度平移信号上的投影，即求 $f(x)$与 $h_{a,b}(x)$的相关性。已有证明，这一子波变换满足能量守恒方程。

3)与加窗傅里叶变换相同，它是冗余的，由于 $a,b$ 是连续变化的，一个分析窗与另一个分析窗大部分内容是重叠的，其相关性很强。

进一步与加窗傅里叶变换相比，有以下特点：

a)从分辨率来看，子波变换较好地解决了时间和频率分辨率的矛盾，它巧妙地利用了非均匀分布的分辨率：在低频段用高的频率分辨率和低的时间分辨率，而在高频段则采用低的频率分辨率和高的时间分辨率。即子波分析的窗宽是可变的，在高频时使用短窗口，而在低频时，则使用宽窗口。这充分体现了常相对带宽对带宽频率分析和自适应分辨分析的思想，它与时变信号特性是一致的，为此用它进行时变信号的分析是十分有效的。

b)从正交性来看，加窗傅里叶变换虽然也是一种时频二维分析，但它与傅里叶变换一样都是正交的：它的正交基底是$\{e^{i\omega x}\}$，即无限长的正弦或余弦，而在频率轴上只是"点频率"，这样对时变信号，由于其频率成份比较丰富，展开系数的能量必然很宽；与此不同，子波变换并不一定要求是正交的，其时宽频宽乘积很小，因而展开系数的能量较为集中。

c)从频谱分析来看，子波变换将信号分解为对数中具有相同大小频带的集合，与加窗傅里叶变换相比，加窗傅里叶变换对不同的频率分量，在时域中都取相同的窗宽，而子波变换窗宽则是可调的。这种以对数形式(非线性的)而不是以线性方式处理频率的方法对时变信号具有明显的优越性，从而可见非线性方法的潜力。

d)从群的观点来看，子波变换与傅氏变换都具有统一性和相似性，其正反变换都具有完美的对称性，因此，我们讲子波分析是傅里叶分析的新发展。

总之，子波展开保留了傅里叶展开的特点，且在时间上和频率上都可进行局域分析，同时由于 $h_{a,b}(x)$是由基本子波函数 $h(x)$(或称为母波)经平移和伸缩变换构造的。因此频谱分析仍可进行，只是基波 $e^{ix}$须用 $h(x)$代替。

### 三、子波变换的应用简例

正是由于子波变换既保持了傅里叶分析的优点，又弥补了傅里叶分析的不足，尤其是它对时变的非平稳信号的独特处理技术，使它在众多的科学技术领域受到了高度重视，并得到了广泛应用：从图像处理到语音合成，从地震信号处理到系统故障探测，以至大规模并行科学计算中的快速计算和实时处理中都发挥出了重大的作用。下面列举几个例子简述如下：

1. *子波变换在图像压缩编码中的应用*

子波变换可以把信号分解为多个具有不同的时间和频率分辨率的信号，从而可以在一个

变换中同时研究低频长时现象和高频短时现象。因此用子波变换对图像这种不平稳的复杂信源进行处理时，就能有效地克服用傅里叶分析和其它分析方法所存在的不足。低比特率的数据表示高质量的静止、运动图像是诸如多媒体技术、数字电视、高清晰度要求的一项关键技术。现有的静止、运动图像的压缩国际标准是采用离散余弦变换(DCT)的 JPEG、MPEG－1、MPEG－2，由于 DCT 固有的缺点及分块效应，进一步压缩静止、运动图像极为困难。子波变换由于既有传统 DCT 的正交变换的能量紧致性，又具有子带编码方法的易于控制各带噪声的特性，同时子波变换还具有与人类视觉系统相吻合的对数特征，因此子波变换在图像压缩领域受到密切关注，现已成为获得低比特率高质量图像的一个重要方法。

Mallat 第一个将子波变换应用于图像压缩，接着 Antonini[2-6]等用双正交子波变换方法完成了一个使用多精度码本的矢量量化方案，并指出子波变换特别适用于图像累进传输的要求；Desarte 等给出了一种信号的自适应、多精度变换方案。为此专家们预计高图像质量、高压缩比和低比特率传输速率的视频压缩技术将采用子波变换算法。

2. 子波变换在突变信号检测中的应用

信号中的突变部分往往携带着重要的、丰富的信息，如语音中辅音至元音的急剧过渡对于发音的听辨起着重要的作用；而图像中的不连续处通常表现为物体的轮廓等。因此突变信号的检测始终是信号处理中的重要内容之一。子波变换由于其分辨率随频率的不同而变化的特点能对检测的信号提供比较准确的定位，尤其是 Mallat 子波，能准确地定位信号的上升沿和下降沿，对噪声的谱分布不敏感。

有关子波变换在信号检测以及数据压缩的内容请参看参考文献[2－7]。

3. 子波变换在信号重构中的应用

通常使用的采样定理以及信号重构问题都是从傅里叶分析的观点出发的，因此处理的对象都是带限信号。信号的过零包涵了非常丰富的信息。生理学研究表明人类的视觉系统和听觉系统的部分信息的感知是通过检测过零来获取的，过零率检测已被应用于语音信号的端点检测，图像信号的过零率也被作为边沿检测的基础，同时用过零表示信号还是数据压缩的一个重要研究方向。Mallat 从子波变换的角度出发，指出了信号的子波系数实际上是信号经带通滤波的结果，因而在一定的条件下可以从子波系数的过零点或极大(小)值恢复信号，虽然没有从理论上证明子波系数过零唯一确定子波系数，但多数信号的子波系数可以用子波系数的过零来确定，并由此引出用子波变换的方法进行信号的奇异值分析及去噪处理。

4. 子波变换在暂态信号检测、宽带信号处理中的应用

传统的信号检测方法是基于相关运算的匹配滤波器，这就要求精确地知道被检测信号的形状、到达时刻等先验知识，当不能获得这些先验知识时，匹配滤波器的性能将急剧下降。Friedlander 等人最先提出了用 Gabor 变换来得到暂态信息的时频信息，从而达到有效地表示暂态信号，提高检测概率的目的。继 Friedlander 之后，Lee 及 Friedlander 等人又相继提出用过采样的 Gabor 变换进一步来提高检测器的鲁棒性，用时频变换方法进行暂态信号的检测受到研究者的高度重视。子波变换由于具有可变的时频分辨能力，子波变换比 Gabor 变换具有更为灵活的时频表示，Frisch 针对已知信号波形、未知尺度因子和到达时刻及完全未知暂态信号先验知识等情况，使用子波变换方法得到暂态信号的时频信息，获得了满意的效果。在宽带信号处理方面，Wiss 注意到了宽带模糊函数与子波变换的相似性，提出了用子波变换的方法研究宽带信号处理；Jouny 提出了用子波变换来分析雷达目标回波模型数据，并从简化的回波模

型中估计目标散射中心的散射机理。

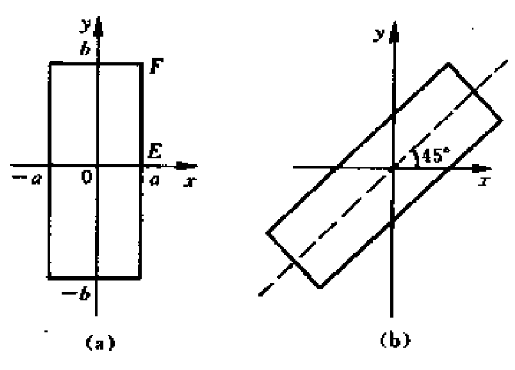
## 习　题

2—1　求下列图像的二维傅里叶变换

(a)长方形图像

$$f(x,y)=\begin{cases}E, & |x|<a; \quad F, \quad |y|<b \\ 0 & \text{其它}\end{cases}$$

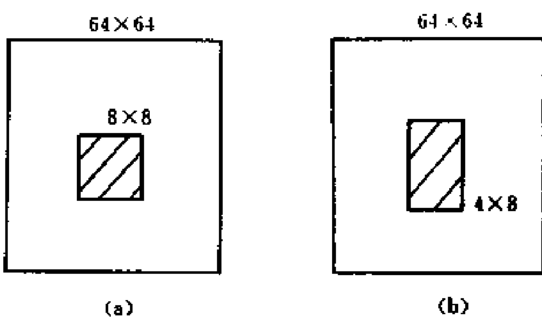
(b)旋转45°后的长方形图像




题2—1图

2—2　写出二维离散傅里叶变换(DFT)对的矩阵表达式及表达式中各个矩阵的具体内容，并以 $N=4$ 为例证明可以从傅里叶正变换矩阵表达式推出反变换的矩阵表达式。

2—3　请实际编程做出以下图像的二维DFT




题2—3图

请打印出以上图像经过变换以后的幅度谱图像(包括数字图像及黑白图像)。

2—4　试编制题2—3中图 $a,b$ 的离散余弦变换的程序，并打印出其幅度谱图像。

## 参 考 文 献


[2—1]　Pratt W.K. Digital Image Processing,1978

[2—2]　Rosenfeld A. Digital Picture Processing,second edition 1982

[2—3]　Gonazales R.C. and Winty P. Digital Image Processing,1977

[2—4]　冈萨雷斯 R.C. 数字图像处理、李叔梁等译。北京:北京科学出版社,1983

[2—5]　李世雄,刘家琦编著,小波变换和反演数学基础。北京:地质出版社,1994

[2—6]　Antonini M. Borland M. Mathien P. and Daubieehes I. Image Coding using wavele transform,IEEE Trans. On Image Processing,Vol. 1 No. 2,Apr. 1992,PP. 205～220

# 第三章　图像的增强

在图像的生成、传输或变换的过程中，由于多种因素的影响，总要造成图像的降质。图像增强的目的是采用一系列技术去改善图像的效果或将图像转换成一种更适合于人或机器进行分析处理的形式。图像增强处理有两类方法：空间域法和频率域法，前者是在原图像上直接进行数据运算，后者是在图像的变换域上进行修改，增强我们感兴趣的部分。下面将介绍几种常用的技术。

## §3.1　灰度级修正

灰度修正是对图像在空间域进行增强的简单而效果明显的方法，根据图像降质不同的原因以及对图像特征的不同要求而采用不同的修正方法。下面主要介绍灰度变换法和直方图修正法。它们是把原图像的灰度函数 $f(x,y)$ 经过一个变换函数 $T(\cdot)$ 变换成一个新的图像函数 $g(x,y)$，即

$$g(x,y) = T[f(x,y)] \tag{3.1.1}$$

通过变换，达到对比度增强的效果，要注意在变换的过程中，对每一个象点 $(x)$ 都经过了同样的处理，因此以上的方法又叫做点处理。

### 一、灰度变换法

一般成象系统只具有一定的亮度范围，亮度的最大值与最小值之比称为对比度。由于形成图像的系统亮度有限，常出现对比度不足的弊病，使人眼观看图像时视觉效果很差。通过下面介绍的灰度变换法可以大大改善人的视觉效果。灰度变换法又可分为以下三种：线性、分段线性以及非线性变换。

1. 线性变换

假定原图像 $f(x,y)$ 的灰度范围为 $[a,b]$，希望变换后图像 $g(x,y)$ 的灰度范围扩展至 $[c,d]$，线性变换表示式为

$$g(x,y)=[(d-c)/(b-a)]f(x,y)+c \tag{3.1.2}$$

此关系式可用图 3－1 表示。若图像中大部分象素的灰度级分布在区间 $[a,b]$，很小部分的灰度级超出了此区间，为改善增强的效果，可令



图 3－1　灰度范围线性变换

$$g(x,y)=\begin{cases} c & 0<f(x,y)<a \\ \dfrac{d-c}{b-a}f(x,y)+c & a\leqslant f(x,y)\leqslant b \\ d & b\leqslant f(x,y)\leqslant M_f \end{cases} \tag{3.1.3}$$

2. 分段线性变换

为了突出感兴趣的目标或灰度区间，相对抑制那些不感兴趣的灰度区域，常采用分段线性

法。常用的是三段线性变换法，如图 3－2 所示，其数学表达式

$$g(x,y)=\begin{cases}(c/a)f(x,y) & 0\leqslant f(x,y)\leqslant a\\ [(d-c)/(b-a)]f(x,y)+c & a\leqslant f(x,y)\leqslant b\\ [(M_g-d)/(M_f-b)][f(x,y)-b]+d & b\leqslant f(x,y)\leqslant M_f\end{cases}\quad(3.1.4)$$

图中对灰度区间$[a,b]$进行了线性变换，而灰度区间$[0,a]$和$[b,M_f]$受到了压缩。通过细心调整折线拐点的位置及控制分段直线的斜率，可对任一灰度区间进行扩展或压缩。这种变换适用于在黑色或白色附近有噪声干扰的情况。例如照片中的划痕，由于变换后 $0\sim a$ 以及 $b\sim M_f$ 之间的灰度受到压缩，因而使污斑得到减弱。

图 3－2　分段线性变换

3．非线性灰度变换

当用某些非线性函数，例如对数函数等作为(3.1.1)式的映射函数时，可实现图像灰度的非线性变换，对数变换的一般式为

$$g(x,y)=a+\frac{\ln[f(x,y)+1]}{b\cdot\ln c}\quad(3.1.5)$$

这里 $a$、$b$、$c$ 是便于调整曲线的位置和形状而引入的参数，它使低灰度范围的 $f$ 得以扩展而高灰度 $f$ 得到压缩，以使图像的分布均匀与人的视觉特性相匹配。

指数变换的一般式为：

$$g(x,y)=b^{c[f(x,y)-a]}-1\quad(3.1.6)$$

其中 $a$,$b$,$c$ 三个参数用来调整曲线的位置和形状，它的效果与对数相反，它将对图像的高灰度区给予较大的扩展。

## 二、直方图修正法

1．概述

直方图表示数字图像中每一灰度级与其出现的频数(该灰度级的象数总数)或相对频数(该灰度级象数与图像的总象数之比)之间的关系。按定义直方图可表示为

$$P(r_k)=n_k/N\quad(3.1.7)$$

式中 $N$ 为一幅图像的总象素数，$n_k$ 为第 $k$ 级灰度的象素数，$r_k$ 表示第 $k$ 个灰度级，$P(r_k)$表示该灰度级出现的相对频数。

一幅图像对应一个直方图，但一个直方图并不一定只对应一幅图像。几幅图像只要灰度分布密度相同，那么它们的直方图也是相同的，假定有一个只有两个灰度级，且分布规律相同的直方图如图 3－3(a)所示，其相对应的图像可以为图 3－3(b)所示的几种不同的图像。

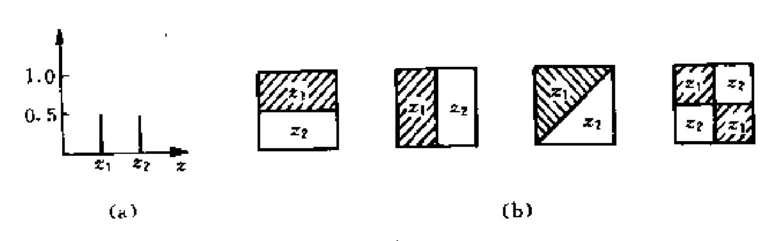



图 3－3　直方图

(a)二值直方图；(b)几个具有相同直方图的图像

2. 直方图均匀化

直方图均匀化是另一类灰度增强的算法。一幅对比度较小的图像，其直方图分布一定集中在某一比较小的范围之内，经过均匀化处理的图像，其所有灰度级出现的相对频数（概率）相同，此时图像的熵最大，图像所包含的信息量最大。图 3-4 所示为连续情况下非均匀概率密度函数 $P_r(r)$ 经变换函数 $T(r)$ 转换为均匀概率分布 $P_s(s)$ 的情况，图中 $r$ 为变换前的归一化灰度级，$0\leqslant r\leqslant 1$。$T(r)$ 为变换函数。$S=T(r)$ 为变换后的灰度值，也归一化为 $0\leqslant S\leqslant 1$。假定(1)在 $0\leqslant r\leqslant 1$ 区间，$T(r)$ 为单调递增函数，且满足 $0\leqslant T(r)\leqslant 1$。(2)反变换 $r=T^{-1}(s)$ 存在，$0\leqslant s\leqslant 1$，也满足类似(1)的条件，变换后的图像灰度范围与原图一致。

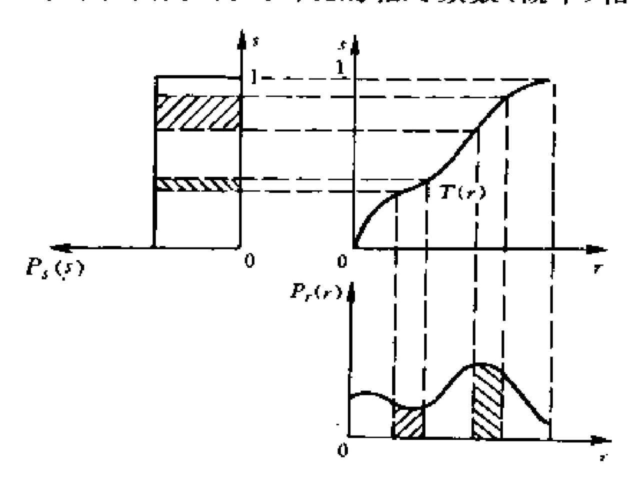


图 3-4　将非均匀密度变换为均匀密度

由概率知识得：

$$P_s(s)=P_r(r)\frac{\mathrm{d}r}{\mathrm{d}s}\bigg|_{r=T^{-1}(s)} \qquad (3.1.8)$$

当直方图均衡化（并归一化）后有 $P_s(s)=1$，即

$$\mathrm{d}s=P_r(r)\cdot\mathrm{d}r=\mathrm{d}T(r)$$

两边取积分得

$$S=T(r)=\int_0^r P_r(r)\mathrm{d}r \qquad (3.1.9)$$

式(3.1.9)就是所求的变换函数，它表明变换函数是原图像的累计分布函数，是一个非负的递增函数。

例 3－1　给定一幅图像的灰度级概率密度函数为

$$P_r(r)=\begin{cases}-2r+2 & 0\leqslant r\leqslant 1\\ 0 & \text{其它}\end{cases}$$

要求其直方图均匀化，计算出变换函数 $T(r)$：

$$S=T(r)=\int_0^r P_r(r)\mathrm{d}r$$
$$=\int_0^r(-2r+2)\mathrm{d}r=-r^2+2r$$

有了 $T(r)$，即可由 $r$ 计算 $S$，亦即由 $P_r(r)$ 分布的图像得到 $P_r(s)$ 的分布图像，至于按 $T(r)$ 变换后的图像灰度级分布是否均匀，即 $P_s(s)$ 是否等于 1，请读者通过计算 $T^{-1}(s)$ 来验证。

在离散的情况下，对总象数为 $N$，$L$ 个灰度级，第 $k$ 个灰度级 $r_k$ 出现的频数为 $n_k$，则第 $k$ 个灰度级出现的概率为

$$P_r(r_k)=n_k/N \qquad (0\leqslant r_k\leqslant 1,k=0,1,\cdots,L-1) \qquad (3.1.10)$$

对其进行均匀化处理的变换函数为

$$S_k=T(r_k)=\sum_{j=0}^{k}P_r(r_j)=\sum_{j=0}^{k}n_j/N \qquad (3.1.11)$$

而逆变换函数为：

$$r_k=T^{-1}(s_k) \qquad 0\leqslant S_k\leqslant 1 \qquad (3.1.12)$$

例 3—2 假设有一幅图像，共有 64×64 个象素，8 个灰度级，各灰度级概率分布见表 3-1(a)，试将其直方图均匀化。

**表 3-1(a) 各灰度级概率分布($N=4096$)**

| 灰度级 $r_k$ | $r_0=0$ | $r_1=1/7$ | $r_2=2/7$ | $r_3=3/7$ | $r_4=4/7$ | $r_5=5/7$ | $r_6=6/7$ | $r_7=1$ |
|---|---|---|---|---|---|---|---|---|
| 象素数 $n_k$ | 790 | 1023 | 850 | 656 | 329 | 245 | 122 | 81 |
| 概率 $P_r(r_k)$ | 0.19 | 0.25 | 0.21 | 0.16 | 0.08 | 0.06 | 0.03 | 0.02 |

**表 3-1(b) 直方图均匀化过程**

| 原灰度级 | 变换函数 $T(r_k)$值 | 象素数 | 量化级 | 新灰度级 | 新灰度级分布 |
|---|---|---|---|---|---|
| $r_0=0$ | $T(r_0)=s_0=0.19$ | 790 | 0 | | 0 |
| $r_1=1/7$ | $T(r_1)=s_1=0.44$ | 1023 | 1/7=0.14 | $S_0'(790)$ | 790/4096=0.19 |
| $r_2=2/7$ | $T(r_2)=s_2=0.65$ | 850 | 2/7=0.29 | | |
| $r_3=3/7$ | $T(r_3)=s_3=0.81$ | 656 | 3/7=0.43 | $S_1'(1023)$ | 1023/4096=0.25 |
| $r_4=4/7$ | $T(r_4)=s_4=0.89$ | 329 | 4/7=0.57 | | |
| $r_5=5/7$ | $T(r_5)=s_5=0.95$ | 245 | 5/7=0.71 | $S_2'(850)$ | 850/4096=0.21 |
| $r_6=6/7$ | $T(r_6)=s_6=0.98$ | 122 | 6/7=0.86 | $S_3'(985)$ | 985/4096=0.24 |
| $r_7=7/7$ | $T(r_7)=s_7=1$ | 81 | 1 | $S_4'(448)$ | 448/4096=0.11 |

现将图像直方图均匀化过程扼要说明如下：根据表 3—1(a)作出的此图像直方图如图 3-5(a)所示，应用式(3—11)可求得变换函数为

$$S_0=T(r_0)=\sum_{j=0}^{0}P_r(r_j)=0.19$$

$$S_1=T(r_1)=\sum_{j=0}^{1}P_r(r_j)=P_r(r_0)+P_r(r_1)=0.19+0.25=0.44$$

按此同样的方法计算出 $S_2,S_3,S_4,S_5,S_6,S_7$ 如下：

$S_2=0.65$　　　$S_5=0.95$

$S_3=0.81$　　　$S_6=0.98$

$S_4=0.89$　　　$S_7=1.00$

图 3-5(c)示出了 $S_K$ 与 $r_k$ 之间的关系曲线，根据变换函数 $T(r_k)$我们可以逐个将 $r_k$ 变成 $S_k$，从表 3-1(a)中看出原图像给定的 $r_k$ 是等间隔的(每个间隔 1/7)，而经过 $T(r_k)$求得的 $S_k$ 就不一定再是等间隔的，从图 3-5(b)中可以明显看出这一点，为了得到等间隔的 $S_k$，必须对每一个变换的 $S_k$ 取最靠近的量化值，表 3-1(b)中列出了重新量化后得到的新灰度 $S_0',S_1',S_2',S_3',S_4'$。

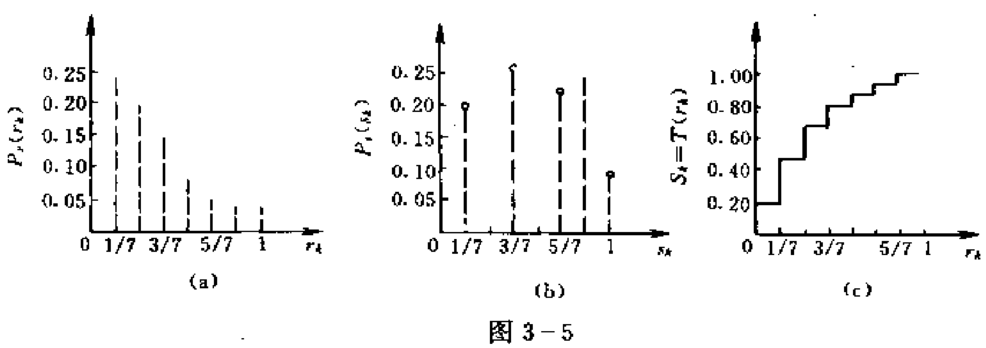



图 3-5

把计算出来的 $S_k$ 与量化级数相比较，可以得出：

$$S_0=0.19\rightarrow\frac{1}{7}\quad S_1=0.44\rightarrow\frac{3}{7}\quad S_2=0.65\rightarrow\frac{5}{7}\quad S_3=0.81\rightarrow\frac{6}{7}$$

$$S_4=0.89\rightarrow\frac{6}{7}\quad S_5=0.95\rightarrow 1\quad S_7=1\rightarrow 1$$

由上可知，经过变换后的灰度级不需要 8 个，而只需要 5 个就可以了，它们是：

$$S'_0=\frac{1}{7}\quad S'_1=\frac{3}{7}\quad S'_2=\frac{5}{7}\quad S'_3=\frac{6}{7}\quad S'_4=1$$

把相应原灰度级的象素数相加就得到新灰度级的象素数。均匀化以后的直方图示于图 3－5(b)，从图中可以看出均匀化后的直方图比原直方图 3－5(a)均衡了，但它并不能完全均匀，这是由于在均衡化的过程中，原直方图上有几个象素数较少的灰度级归并到一个新的灰度级上，而象素较多的灰度级间隔被拉大了。这样减少图像的灰度等级以换取对比度的扩大。如果读者对被压缩部分的图像细节比较感兴趣，则可以采用局部自适应直方图均衡比较适宜，简称为 LAHE(Local Adaptive Histogram Equalization)。

如果希望得到一个直方图完全平均而且灰度等级又不减少的均衡化处理，则必须用一些拟合技术，下面举例说明这种技术。以上述图像为例，要求处理后每一灰度出现的概率相等，则每一灰度的象素必须都是 512，如表 3－2 所示。

**表 3－2　均匀处理结果**

| | $S_0$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | 处理前 |
|---|---|---|---|---|---|---|---|---|---|
| $r_0$ | 512 | 278 | | | | | | | 790 |
| $r_1$ | | 234 | 512 | 277 | | | | | 1023 |
| $r_2$ | | | | 235 | 512 | 103 | | | 850 |
| $r_3$ | | | | | | 409 | 247 | | 656 |
| $r_4$ | | | | | | | 265 | 64 | 329 |
| $r_5$ | | | | | | | | 245 | 245 |
| $r_6$ | | | | | | | | 122 | 122 |
| $r_7$ | | | | | | | | 81 | 88 |
| 处理后 | 512 | 512 | 512 | 512 | 512 | 512 | 512 | 512 | 4096 |

从上表可以看到为了得到均匀输出，原灰度为 $r_0 \sim r_4$ 的象素都必须分成两部分和三部分。例如某一象素点原来的灰度值为 $r_1$，变换后它可以属于 $S_1$、$S_2$ 或 $S_3$。如何确定其确切的灰度值，有两个办法。

(1)随机取数法：以上例的 $r_3$ 为例，原图中属于 $r_3$ 的有 656 个象素。按表上所列其中 409 个应归入 $S_5$，247 个应归入 $S_6$。为确定某个灰度为 $r_3$ 的象素应归入 $S_5$ 还是 $S_6$，用计算机产生一个值为 0～1 间均匀分布的随机数，每遇到一个 $r_3$，就让随机数发生器产生一个随机数。由于 409/656＝0.623；247/656＝0.377，因此，如产生的随机数在 0～0.623 之间就把这一灰度为 $r_3$ 的象素转换成 $S_5$；如随机数在 0.623～1 之间就把这个灰度为 $r_3$ 的象素转为 $S_6$。

(2)按这一象素的邻域点的灰度来决定这个象素应属于哪个输出级。例如对灰度为 $r_3$ 的象素，要看这个象素的邻域点(四邻域或八邻域)有没有 $S_5$、$S_6$ 的象素，以具有 $S_5$ 和 $S_6$ 灰度的象素多少来确定 $r_3$ 属于 $S_5$ 或 $S_6$；如 $r_3$ 的邻域没有 $S_5$ 或 $S_6$ 的，则看其邻域有没有 $S_4$ 或 $S_7$ 的，以确定其输出属于 $S_5$ 或 $S_6$，邻域中 $S_4$ 多则 $r_3$ 属于 $S_5$，$S_7$ 多则属于 $S_6$。

3. 直方图规定化

以上均匀化处理后的图像虽然增强了图像的对比度，但它并不一定适合人的视觉，为此提出直方图规定化处理。

直方图规定化方法之一是用一个规定的概率密度函数来表示直方图，如表 3－3 所示。

表 3－3　几种给定形状的直方图修正变换函数

| | 修正后要求的概率密度函数 $P_s(s)$ | 变换函数 $S=T(r)$ |
|---|---|---|
| 均匀分布 | $P_s=1/S_{max}-S_{min}$<br>$S_{min}\leqslant S\leqslant S_{max}$ | $S=[S_{max}-S_{min}]P_r(r)+S_{min}$ |
| 指数分布 | $P_s(s)=\alpha\exp[-\alpha(S-S_{min})]$<br>$S\geqslant S_{min}$ | $S=S_{min}-\frac{1}{\alpha}\ln[1-P_r(r)]$ |
| 雷利分布 | $P_s(s)=\frac{S-S_{min}}{\alpha^2}\exp\left\{\frac{-(S-S_{min})^2}{2\alpha^2}\right\}$ | $S=S_{min}+[2\alpha^2\ln(\frac{1}{1-P_r(r)})]^{1/2}$ |
| 双曲线 | $P_s(s)=\frac{1}{S[\ln(S_{max})-\ln(S_{min})]}$ | $S=S_{min}[\frac{S_{max}}{S_{min}}]^{P_r(r)}$ |

一般先将函数数学化，再形成直方图。具体见参考文献[3－1]。

规定化处理的第二种方法是通过控制一组直线段来构成直方图，使其满足希望的形状。然后再数字化并归一化，图 3－6 中的直线段构成直方图形状受 $m,h,\theta_L,\theta_R$ 四个参量控制，其中 $m$ 在[0,1]区间内任意选定；$h>0$；$\theta_L,\theta_R$ 从 0°到 90°变化，分别引起 $j,k$ 点沿两点[(0,1)和$(m,0)$，(1,1)和$(m,0)$]的连线上移动，只要改变上述四个参量就可以得到许多有用的直方图。如果取 $m=0.5,h=1,\theta_L=\theta_R=0$ 就可以得到一个矩形，即均匀直方图。

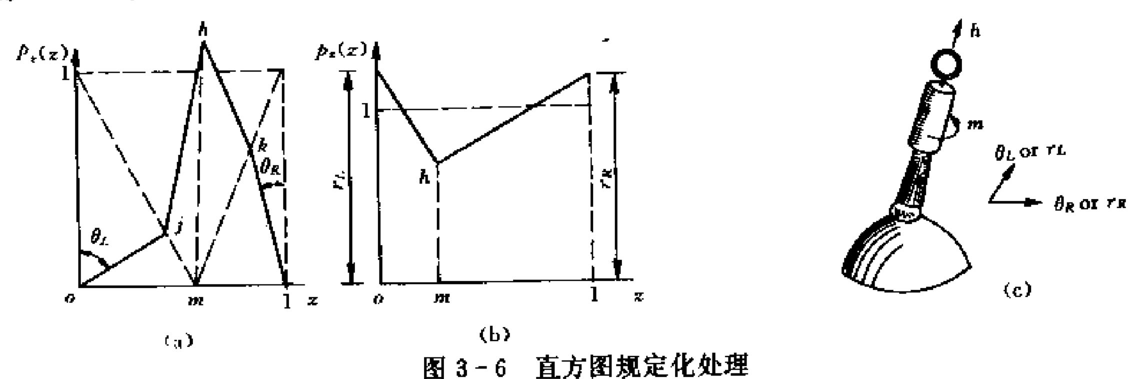



图 3－6　直方图规定化处理

下面具体讨论如何实现直方图规定化处理。先讨论连续的情况：设 $P_r(r)$和 $P_z(z)$分别代表原始图像和规定化处理后的灰度概率密度函数，分别对原始直方图和规定化处理后的直方图进行均匀化处理，即

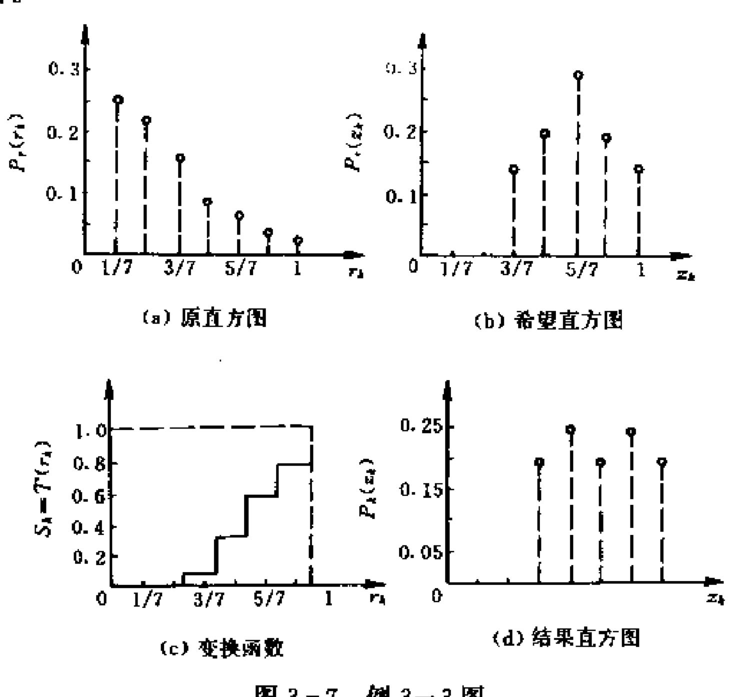
$$S=T(r)=\int_0^r P_r(w)\mathrm{d}w \tag{3.1.13}$$

$$V=G(z)=\int_0^Z P_Z(w)\mathrm{d}w \tag{3.1.14}$$

均匀化处理后的二者所获得的图像灰度概率密度函数 $P_s(s)$和 $P_v(v)$应该是相等的(均为 1)，为此可以用 $s$ 代替 $v$，再使用 $Z=G^{-1}(v)=G^{-1}(s)$逆变换来获得希望的图像。

例 3－3　仍用例 3－2 的原始直方图，其规定化后的直方图如图 3－7(b)所示，把计算过

程由下表 3-4 列出。




(a) 原直方图　(b) 希望直方图

(c) 变换函数　(d) 结果直方图

图 3-7　例 3—3 图

**表 3-4　例 3—3 运算过程**

| ① | ② | ③ | ④ |
|---|---|---|---|
| $r_0 \to S_0^1=1/7=0.14$ | $Z_0-V_0=0$ | | |
| $r_1 \to S_1^1=3/7=0.43$ | $Z_1-V_1=0$ | | |
| $r_2 \to S_2^1=5/7=0.71$ | $Z_2-V_2=0$ | | |
| $\left.\begin{matrix} r_3 \\ r_4 \end{matrix}\right\} \to S_3^1=6/7=0.86$ | $Z_3-V_3=0.15$ | $\to S_0^1 \to Z_3^1$ | $\xrightarrow{(790)} 0.19$ |
| | $Z_4-V_4=0.35$ | $\to S_1^1 \to Z_4^1$ | $\xrightarrow{(1023)} 0.25$ |
| $\left.\begin{matrix} r_5 \\ r_6 \\ r_7 \end{matrix}\right\} \to S_4^1=1$ | $Z_5-V_5=0.65$ | $\to S_2^1 \to Z_5^1$ | $\xrightarrow{(850)} 0.21$ |
| | $Z_6-V_6=0.85$ | $\to S_3^1 \to Z_6^1$ | $\xrightarrow{(985)} 0.24$ |
| | $Z_7-V_7=1$ | $\to S_4^1 \to Z_7^1$ | $\xrightarrow{(448)} 0.11$ |

以上四个步骤如下：

①重复例 3—2 的均匀化过程，8 个尺度级并为 5 个尺度级

②对规定化的图像用同样的方法进行直方图均匀化处理，$V_k=G(z_k)=\sum_{j=0}^{k}P_z(z_j)$

③使用与 $V_k$ 靠近的 $S_k$ 代替 $V_k$，并用 $G^{-1}(s)$ 求逆变换即可得到 $z_k'$。

④图像总象素点为 $64\times64=4096$，根据一系列 $z_k^1$ 求出相应的 $P_k(z_k)$。其结果如图 3-7(d)。

### 三、图像的同态增晰

图像的同态增晰法属于图像频率域处理范畴，而它的作用仍是对图像的灰度范围进行调整。因此把它作为灰度修正的一种方法放在此处叙述。

往往会得到这样的图像，它的动态范围很大，而我们感兴趣的部分的灰度级又很少，图像

的细节没办法辨认，采用一般的灰度级线性变换法是不行的，为此可采用同态增晰法，此法的框图如图 3－8(a)所示。

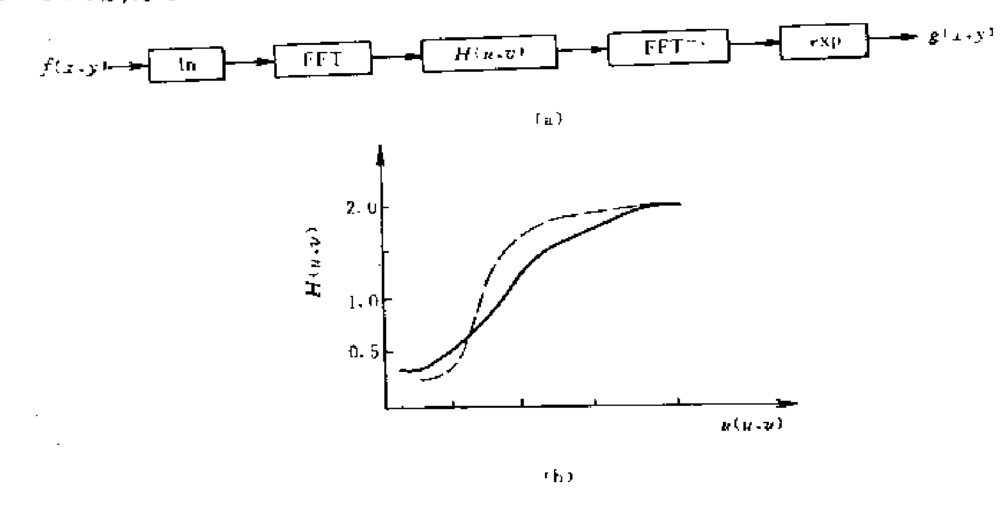



图 3－8　同态增晰原理框图

一般自然景物的图像 $f(x,y)$ 可以由照明函数 $f_i(x,y)$ 和反射函数 $f_r(x,y)$ 的乘积来表示，照明函数 $f_i(x,y)$ 描述景物的照明，与景物无关；反射函数包含景物的细节，与照明无关。一般 $f_i(x,y)$ 是有限的，而反射函数是小于 1 的，且均为正值。

$$f(x,y)=f_i(x,y)\cdot f_r(x,y) \tag{3.1.15}$$

$$0<f_i(x,y)<\infty \qquad 0<f_r(x,y)<1 \tag{3.1.16}$$

(1)首先对式(3.1.15)取对数

$$\begin{aligned}\ln f(x,y)&=\ln[f_i(x,y)\cdot f_r(x,y)]\\&=\ln f_i(x,y)+\ln f_r(x,y)\end{aligned} \tag{3.1.17}$$

(2)对上式取傅里叶变换，得

$$\begin{aligned}F_{\ln}(u,v)&=F[\ln f_i(x,y)+\ln f_r(x,y)]\\&=F_{i,\ln}(u,v)+F_{r,\ln}(u,v)\end{aligned} \tag{3.1.18}$$

由于场景的照明亮度一般是缓慢变化的，所以照明函数的频谱特性集中在低频段，而景物本身具有较多的细节和边缘，为此反射函数的频谱集中在高频段。另一方面，照明函数描述的图形分量变化幅度大而包含的信息少，而反射函数描述的景物，图像的灰度级较少而信息较多，为此必须将其扩展。

(3)将式(3.1.18)乘上 $H(u,v)$，其特性见图 3－8。很明显，低频段被压缩，而高频段却扩展了。

$$\begin{aligned}G_{\ln}(u,v)&=F_{\ln}(u,v)\cdot H(u,v)\\&=F_{i,\ln}(u,v)\cdot H(u,v)+F_{r,\ln}(u,v)\cdot H(u,v)\\&=G_{i,\ln}(u,v)+G_{r,\ln}(u,v)\end{aligned} \tag{3.1.19}$$

(4)求傅里叶反变换以及求指数，得

$$\begin{aligned}F^{-1}\{G_{\ln}(u,v)\}&=\ln g_i(x,y)+\ln g_r(x,y)\\&=\ln[g_i(x,y)\cdot g_r(x,y)]\end{aligned} \tag{3.1.20}$$

$$\begin{aligned}g(x,y)&=\exp\{\ln g_i(x,y)\cdot g_r(x,y)\}\\&=g_i(x,y)\cdot g_r(x,y)\end{aligned} \tag{3.1.21}$$

根据不同的图像特性和需要，选用不同的 $H(u,v)$，可以得到满意的结果。要注意如果图像信号是一个复数，如全息图像，则还有一个多值问题。关于同态处理请看参考文献[3—2]。

# §3.2 图像的平滑

## 一、引言

图像平滑的目的是为了消除噪声。图像噪声的来源有三：一为在光电、电磁转换过程中引入的人为噪声；二为大气层电(磁)暴、闪电、电压、浪涌等引起的强脉冲性冲激噪声的干扰；三为自然起伏性噪声，由物理量的不连续性或粒子性所引起，这类噪声又可分成热噪声、散粒噪声等。噪声消除的方法又可以分为空间域或频率域，亦可以分为全局处理或局部处理，亦可以按线性平滑、非线性平滑和自适应平滑来区别。下面介绍邻域平均、空域滤波、频域滤波、多图像平均、自适应滤波及中值滤波等方法。

## 二、邻域平均法

邻域平均法是一种局部空间域处理的算法。设一幅图像 $f(x,y)$ 为 $N\times N$ 的阵列，平滑后的图像为 $g(x,y)$，它的每个象素的灰度级由包含在 $(x,y)$ 的预定邻域的几个象素的灰度级的平均值所决定，即用下式得到平滑的图像。

$$g(x,y)=\frac{1}{M}\sum_{(i,j)\in S}f(i,j) \tag{3.2.1}$$

式中的 $x,y=0,1,2,\cdots,N-1$，$S$ 是 $(x,y)$ 点邻域中心点的坐标的集合[不包括点 $(x,y)$]，$M$ 是 $S$ 内坐标点的总数。图 3-9 示出了四个邻域点和八个邻域点的集合。图(a)中的邻域半径为一个阵元间隔 $\Delta x$，图(b)的邻域半径为 $\sqrt{2}\Delta x$ 的间隔。

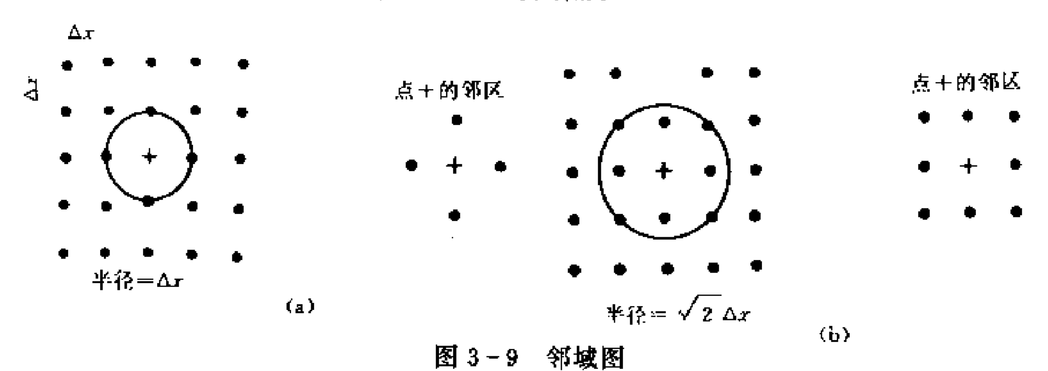



图 3-9 邻域图

图 3-10 中示出了用邻域平均法所产生的平滑效果，由图可知，所用的邻域半径愈大，则图的模糊程度越大。

以上算法简单，计算速度快，但它的主要缺点是在降低噪声的同时使图像产生模糊，特别在边沿和细节处，邻域越大，模糊越厉害。

为了减少这种效应，可以采用阈值法，亦就是根据下列准则形成平滑图像

$$g(x,y)=\begin{cases}\dfrac{1}{M}\sum\limits_{(m,n)\in S}f(m,n), & 若\left|f(x,y)-\dfrac{1}{M}\sum\limits_{(m,n)\in S}f(m,n)\right|>T\\ f(x,y) & 其它\end{cases} \tag{3.2.2}$$

式中 $T$ 是一个规定的非负阈值，当一些点和它们邻值的差值不超过规定的 $T$ 阈值时，仍保留

这些点的象素灰度值。这样平滑后的图像比邻域平均法模糊度减少。当某些点的灰度值与各邻点灰度的均值差别较大时，它必然是噪声，则取其邻域平均值作为该点的灰度值，它的平滑效果仍然是很好的。

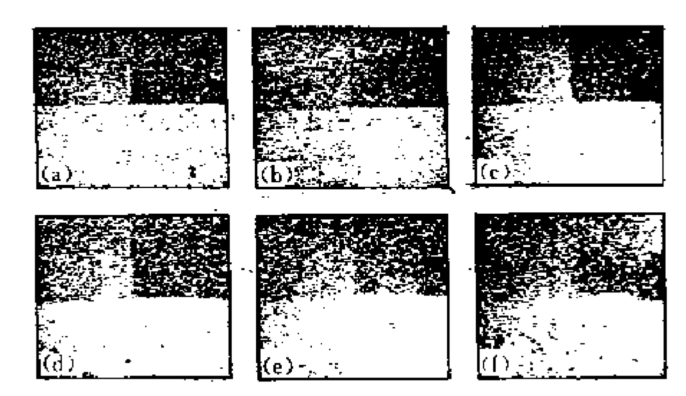



图 3-10　邻域平均法实例
(a)噪声图像；
(b)～(f)分别是利用邻域半径为 1、2、4、8 和 16 处理过的图像

为了克服简单局部平均的弊病，目前已提出许多保边沿保细节的局部平滑算法，它们讨论的课题都在如何选择邻域的大小、形状和方向，如何选择参加平均的点数以及邻域各点的权重系数等，它们有：灰度最相近的 $K$ 个邻点平均法，梯度倒数加权平滑，最大均匀性平滑，小斜面模型平滑等等，有关这些方法请参阅参考文献[3－3]。

如果将受噪声干扰的图像看成是一个二维随机场，则可以运用统计理论来分析受噪声干扰的图像平滑后的信噪比问题，一般在噪声属于加性噪声，并且是独立同分布的高斯白噪声(均值为零，方差为 $\alpha^2$)的情况下，我们定义信噪比为含噪图像的均值与噪声方差之比，则含噪图像经邻域平均法平滑之后，其信噪比将提高 $\sqrt{M}$ 倍($M$ 为邻域中包含的象素数目)，可见邻域取得愈大，象点愈多，则信噪比提高愈大，平滑效果好。实际结果见图 3　10。

## 三、空间域低通滤波

从信号频谱分析的知识，我们知道信号的慢变部分在频率域属于低频部分，而信号的快变部分在频率域是高频部分，对图像来说，它的边缘以及噪声干扰的频率分量都处于空间频率域较高的部分，因此可以采用低通滤波的方法来去除噪声，而频域的滤波又很容易从空间域的卷积来实现，为此只要适当地设计空间域系统的单位冲激响应矩阵就可以达到滤除噪声的效果。

$$G(x,y)=\sum_m\sum_n F(m,n)H(x-m+1,y-n+1) \tag{3.2.3}$$

式中：$G$ 为 $N\times N$ 阵列，$H$ 为 $L\times L$ 阵列

下面是几种用于噪声平滑的系统单位冲激响应阵列

$$H_1=\frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}\quad H_2=\frac{1}{10}\begin{bmatrix}1&1&1\\1&2&1\\1&1&1\end{bmatrix}\quad H_3=\frac{1}{16}\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix} \tag{3.2.4}$$

以上矩阵 $H$ 又叫低通卷积模板。

## 四、频域低通滤波法

这是一种频域处理法，对于一幅图像，它的边缘、跳跃部分以及噪声都代表图像的高频分量，而大面积的背景区和慢变部分则代表图像的低频分量，用频域低通滤波法除去其高频分量就能去掉噪声，从而使图像得到平滑。

利用卷积定理，可以把式(3.2.3)写成以下形式

$$G(u,v)=H(u,v)\cdot F(u,v) \tag{3.2.5}$$

式中，$F(u,v)$是含噪图像的傅里叶变换，$G(u,v)$是平滑后图像的傅里叶变换，$H(u,v)$是传递函数。利用 $H(u,v)$使 $F(u,v)$的高频分量得到衰减，得到 $G(u,v)$后再经过反变换就得到所希

望的图像 $g(x,y)$ 了。低通滤波平滑图像的系统框图如图 3－11 所示。

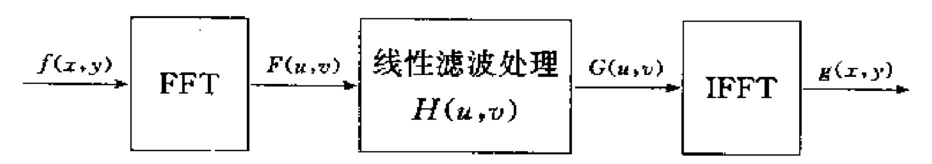



图 3－11 线性滤波处理框图

下面介绍几种常用的低通滤波法：

1. 理想低通滤波器(ILPF)

一个理想的低通滤波器的传递函数由下式表示：

$$H(u,v)=\begin{cases}1 & D(u,v)\leqslant D_0\\ 0 & D(u,v)>D_0\end{cases} \tag{3.2.6}$$

式中 $D_0$ 是一个规定的非负的量，它叫做理想低通滤波器的截止频率。$D(u,v)$代表从频率平面的原点到$(u,v)$点的距离，即：

$$D(u,v)=[u^2+v^2]^{1/2}$$

理想低通滤波器平滑处理的概念是清楚的，但它在处理过程中会产生较严重的模糊和振铃现象。这是由于 $H(u,v)$在 $D_0$ 处由 1 突变到 0，这种理想的 $H(u,v)$对应的冲激响应 $h(x,y)$在空域中表现为同心环的形式，并且此同心环数与 $D_0$ 成反比；$D_0$ 小，同心环数越多，模糊程度愈厉害。正是由于理想低通滤波存在此"振铃"现象，使其平滑效果下降，为此我们将继续介绍指数、梯形和巴特沃思滤波器。

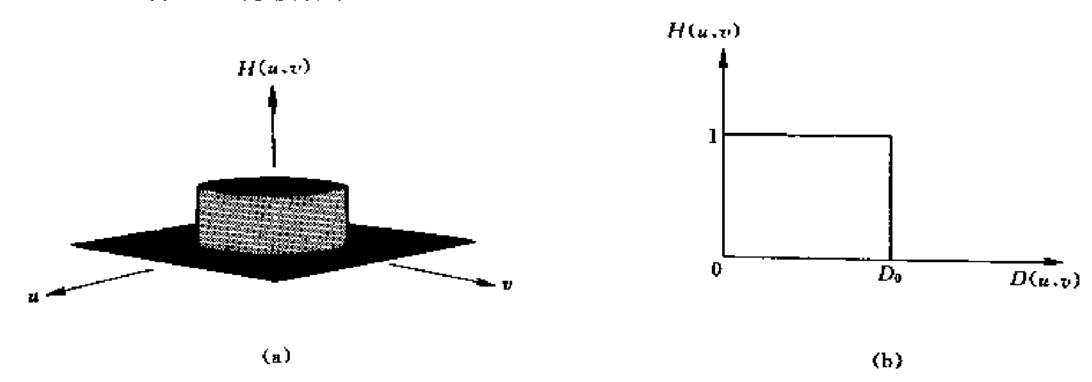



图 3－12 ILPF 特性曲线

2. 巴特沃思低通滤波器(BLPF)

一个 $n$ 阶巴特沃思滤波器的传递系数为

$$H(u,v)=\frac{1}{1+\left[\frac{D(u,v)}{D_0}\right]^{2n}} \tag{3.2.7}$$

或

$$H(u,v)=\frac{1}{1+[\sqrt{2}-1][D(u,v)/D_0]^{2n}} \tag{3.2.8}$$

BLPF 又称作最大平坦滤波器。与 ILPF 不同，它的通带与阻带之间没有明显的不连续性，因此它没有"振铃"现象发生，模糊程度减少，但从它的传递函数特性曲线 $H(u,v)$可以看出，在它的尾部保留有较多的高频，所以对噪声的平滑效果还不如 ILPF。一般情况下，常采用下降到 $H(u,v)$最大值的 $1/\sqrt{2}$ 那一点为滤波器的截止频率点。对式(3.2.8)，当 $D(u,v)=D_0$，$n=1$ 时，$H(u,v)=1/\sqrt{2}$，而对式(3.2.7)$H(u,v)=1/2$，说明两种 $H(u,v)$具有不同的衰减特性，

可以视需要来确定。

3. 指数滤波器(ELPF)

ELPF 的传递函数 $H(u,v)$表示为：

$$H(u,v)=e^{-\left[\frac{D(u,v)}{D_0}\right]^n} \tag{3.2.9}$$

或
$$H(u,v)=\exp^{\left[\ln\frac{1}{\sqrt{2}}\right]\left[\frac{D(u,v)}{D_0}\right]^n} \tag{3.2.10}$$

当 $D(u,v)=D_0,n=1$ 时，对式(3.2.9)，$H(u,v)=1/e$，对式(3.2.10)，$H(u,v)=1/\sqrt{2}$，所以两者的衰减特性仍有不同。由于 ELPF 具有比较平滑的过渡带，为此平滑后的图像没有振铃现象，而 ELPF 与 BLPF 相比，它具有更快的衰减特性见图 3－13，所以经 ELPF 滤波的图像比 BLPF 处理的图像稍微模糊一些。

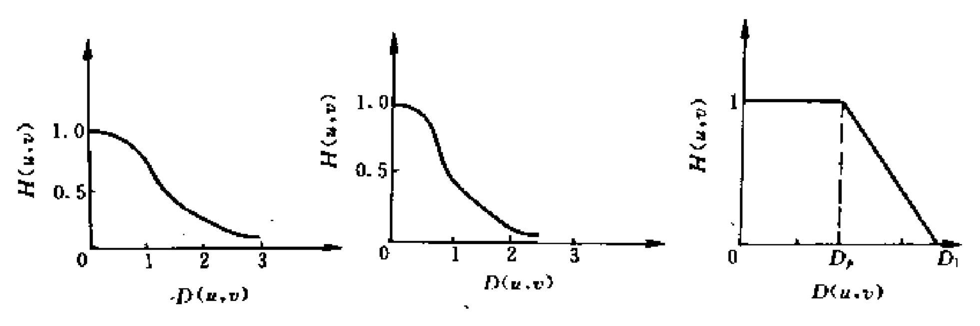



图 3－13　BLPF、ELPF 和 TLPF 特性曲线

4. 梯形滤波器(TLPF)

梯形滤波器的传递函数介于理想低通滤波器和具有平滑过渡带的低通滤波器之间，它的传递函数为：

$$H(u,v)=\begin{cases}1 & D(u,v)<D_0\\ \frac{1}{[D_0-D_1]}[D(u,v)-D_1] & D_0\leqslant D(u,v)\leqslant D_1\\ 0 & D(u,v)>D_1\end{cases} \tag{3.2.11}$$

在规定 $D_0$ 和 $D_1$ 时，要满足 $D_0<D_1$ 的条件。一般为了方便，把 $H(u,v)$的第一个转折点 $D_0$ 定义为截止频率，第二个变量 $D_1$ 可以任意选取，只要 $D_1$ 大于 $D_0$ 就可以了，把四种滤波器的性能列表如下：

| 类　别 | 振铃程度 | 图像模糊程度 | 噪声平滑效果 |
|---|---|---|---|
| ILPF | 严重 | 严重 | 最好 |
| TLPF | 较轻 | 轻 | 好 |
| ELPF | 无 | 较轻 | 一般 |
| BLPF | 无 | 很轻 | 一般 |

## 五、多图像平均法

多幅图像平均法是利用对同一景物的多幅图像取平均来消除噪声产生的高频成分。设原图像为 $f(x,y)$，图像噪声为加性噪声 $n(x,y)$，则下式成立

$$g(x,y)=f(x,y)+n(x,y) \tag{3.2.12}$$

若图像噪声是互不相关的加性噪声，且均值为 0，则 $f(x,y)$为 $g(x,y)$的期望值，取 $M$ 张内容

相同但含有不同噪声的图像，将它们叠加起来，再作平均计算后有

$$\bar{g}(x,y)=\frac{1}{M}\sum_{i=1}^{M}g_i(x,y) \tag{3.2.13}$$

$$f(x,y)=E\{\bar{g}(x,y)\} \tag{3.2.14}$$

$$\sigma^2_{\bar{g}(x,y)}=\frac{1}{M}\sigma^2_n(x,y) \tag{3.2.15}$$

式中 $E\{\quad\}$代表数学期望，$\sigma^2_{\bar{g}(x,y)}$和 $\sigma^2_{n(x,y)}$是 $\bar{g}$ 和 $n$ 在点$(x,y)$处的方差，式(3.2.15)表明对 $M$ 幅图像平均可把噪声方差减少 $M$ 倍，当 $M$ 增大时，$\bar{g}(x,y)$将更加接近于 $f(x,y)$。多图像取平均处理常用于摄象机的进图中，用以减少电视摄象机光导析象管的噪声。这时对同一景物连续摄取多幅图像并数字化，再对多幅图像平均，一般选用 8 幅图像取平均，这种方法的实际应用困难是难于把多幅图像配准起来，以便使相应的象素能正确地对应排列。

## 六、自适应平滑滤波

考虑到图像 $f(x,y)$中目标物体和背景一般都具有不同的统计特性，即有不同的均值和方差。因此，为了保留一定的边缘信息，可采用动态的或自适应的局部平滑滤波，这样当目标物体较大时，可得到较好的图像细节特性。

令

$$\hat{f}(x,y)=\hat{f}(x,y)+k(x,y)\{f(x,y)-\hat{f}(x,y)\} \tag{3.2.16}$$

$k(x,y)$为滤波常数，它随象素位置$(x,y)$而变

$$k(x,y)=\frac{P(x,y)}{P(x,y)+\sigma^2} \tag{3.2.17}$$

$\sigma^2$ 为干扰噪声的方差，这是事先已知的，$P(x,y)$为图像在$(x,y)$点附近的局部统计方差。

$$P(x,y)=\sum_{i=0}^{n}\sum_{j=0}^{n}\frac{1}{N}[f(x+i,y+j)-f(x,y)]^2-\sigma^2 \tag{3.2.18}$$

$N=n\times n$ 为局部均值和局部方差的窗口的大小。

## 七、中值滤波

中值滤波是一种非线性滤波，由于它在实际运算过程中并不需要图像的统计特性，所以比较方便。中值滤波首先是被应用在一维信号处理技术中，后来被二维图像信号处理技术所引用。在一定的条件下，可以克服线性滤波器所带来的图像细节模糊，而且对滤除脉冲干扰及图像扫描噪声最为有效。但是对一些细节多，特别是点、线、尖顶细节多的图像不宜采用中值滤波的方法。

### 1. 中值滤波原理

中值滤波就是用一个含有奇数点的滑动窗口，将窗口正中那点值用窗口内各点的中值代替。假设窗口有 5 点，其值为 80，90，200，110，120，那么此窗口内各点的中值即为 110。

设有一个一维序列 $f_1f_2\cdots f_n$。取窗口长度为 $m$($m$ 为奇数)，对此序列进行中值滤波，就是从输入序列中相继抽出 $m$ 个数，$f_{i-v}\cdots f_{i-1}f_if_{i+1}\cdots f_{i+v}$，其中 $f_i$ 为窗口的中心值，$V=\frac{m-1}{2}$，再将这 $m$ 个点值按其数值大小排列，取其序号为正中间的那个数作为滤波输出。用数学公式表示为

$$Y_i=Med\{f_{i-v}\cdots f_i\cdots f_{i+v}\}\quad i\in Z,\quad V=\frac{m-1}{2} \tag{3.2.19}$$

例如：有一个序列为{0,3,4,0,7}，重新排序后为{0,0,3,4,7}则 $Med\{0,3,4,0,7\}=3$。此例若用平滑滤波，窗口也是取 5，那么平滑滤波输出为(0+3+4+0+7)/5=2.8。与平滑滤波相比，如平滑滤波的窗口也是 $m$，则平滑滤波的输出为 $Z_i=(x_{i-v}+\cdots+x_i+\cdots+x_{i+v})/m \quad i\in Z$就上例而言有

$$Z_i=\frac{0+3+4+0+7}{5}=2.8$$

中值滤波的一个重要特点是可以保持输入波型的上升边缘。

对二维序列$\{X_{ij}\}$进行中值滤波时，滤波窗口也是二维的，但这种二维窗口可以有各种不同的形状，如线状、方形、圆形、十字形、圆环形等。二维数据的中值滤波可以表示为：

$$Y_{ij}=\underset{A}{Med}\{X_{ij}\} \qquad A\text{ 为窗口} \tag{3.2.20}$$

在图像阵列进行中值滤波时，如窗口是以中心点对称的，并包含中心点在内，即

$$(r,s)\in A;(-r,-s)\in A;(0,0)\in A$$

$(r,s)$为窗口内一点与窗口中心的坐标距离，则中值滤波能保持任意方向的跳变边缘。图像中的跳变边缘是指图像中不同灰度区域之间的灰度突变边缘。在实际使用窗口时，窗口的尺寸一般先用 3 再取 5 逐点增大，直到其滤波效果满意为止。对于有缓变的较长轮廓线物体的图像，采用方形或圆形窗口为宜，对于包含尖顶角物体的图像，适宜用十字形窗口。使用二维中值滤波最值得注意的是要保持图像中有效的细线状物体。如果图像中点、线、尖角细节较多，则不宜采用中值滤波。

2. 中值滤波的主要特性

(1)对某些输入信号中值滤波的不变性。对某些特定的输入信号，滤波输出保持输入信号值不变，如在窗口 $2n+1$ 内单调增加或单调减少的序列，即

$$x_{i-n}\leqslant\cdots\leqslant x_i\leqslant\cdots\leqslant x_{i+n} \quad \text{或} \quad x_{i-n}\geqslant\cdots\geqslant x_i\geqslant\cdots\geqslant x_{i+n}$$

则
$$\{y_i\}=\{x_i\}$$

二维序列的中值滤波的不变性要复杂得多，它不但与输入信号有关，而且还与窗口的形状有关。图 3-14 列出了几种二维窗口及与之对应的最小尺寸的不变输入图形。一般讲，与窗口对顶角线垂直的边缘保持不变性。利用这个特点，可以使中值滤波既能去除图像中的噪声，又能保持图像中一些物体的边缘。

图 3-14　中值滤波几种常用窗口及其相应的不变图形

中值滤波的另一类不变性就是在一维情况下周期性的二值序列如：

$$\{x_n\}=\cdots,+1,+1,-1,-1,+1,+1,-1,-1,\cdots$$

若设窗口长度为 9，则中值滤波对此序列保持不变性，也就是说，当窗口为 9 的中值滤波的输入为这一周期为 4 的输入序列时，输出不变。对于一个二维序列，这一类不变性更为复杂，但它们一般也是二值的周期性结构，即周期性网络结构的图像。

(2)中值滤波去噪声性能：中值滤波可以用来减弱随机干扰和脉冲干扰。由于中值滤波是非线性的，因此对随机输入信号数学分析比较复杂。对于均值为零的正态分布的噪声输入，中值滤波输出的噪声方差为

$$\sigma_{Med}^2=\frac{1}{4mf^2(\overline{m})}\approx\frac{\sigma_i^2}{m+\frac{\pi}{2}-1}\cdot\frac{\pi}{2} \tag{3.2.21}$$

式中 $\sigma_i^2$ 为输入噪声功率(方差)，$m$ 为中值滤波窗口长度，$\overline{m}$ 为输入噪声均值；$f(\overline{m})$为输入噪声密度函数。而平滑滤波的输出噪声方差 $\sigma_o^2$ 为

$$\sigma_o^2=\frac{1}{m}\sigma_i^2 \tag{3.2.22}$$

比较式(3.2.21)和式(3.2.22)可看出，中值滤波的输出与输入噪声的密度分布有关。而平均值滤波的输出与输入分布无关。对随机噪声的抑制能力方面来看，中值滤波性能要比平均值滤波差一些。对脉冲干扰来讲，特别是脉冲宽度小于 $m/2$，相距较远的窄脉冲干扰，中值滤波是很有效的。

(3)中值滤波的频谱特性：由于中值滤波是非线性的，为此在输入与输出之间不存在一一对应的关系，故不能用一般线性滤波器频率特性的研究方法。当然我们也可以广义地定义输入与输出之间的关系，即：

$$H=\left|\frac{G}{F}\right|$$

上式中 $G$ 为输入信号频谱，$F$ 为输出信号频谱。可以发现中值滤波的频率特性呈现不规则的波动，而且是与 $G$ 的频谱有关的变动函数，因此对中值滤波不能作频谱分析。但它的 $H$ 曲线波动不大，均值平坦，为此可以认为信号经中值滤波后，频谱基本不变。这点对设计和使用中值滤波都有很大的意义。

3. 复合型中值滤波

对一些内容复杂的图像，可以使用复合型中值滤波。如中值滤波线性组合、高阶中值滤波组合、加权中值滤波以及迭代中值滤波等。

(1)中值滤波的线性组合：将几种窗口尺寸大小和形状不同的中值滤波器复合使用，只要各窗口都与中心对称，滤波输出可保持几个方向上的边缘跳变，而且跳变幅度可调节。其线性组合方程式如下：

$$\overline{Y}_{ij}=\sum_{k=1}^{N}a_k\underset{A_K}{\mathrm{Med}}(f_{ij}) \tag{3.2.23}$$

式中 $a_k$ 为不同中值滤波的系数，$A_K$ 为窗口。

(2)高阶中值滤波组合：可以用式：

$$Y_{ij}=\max_{K}[\underset{A_K}{\mathrm{Med}}(x_{ij})] \tag{3.2.24}$$

这种中值滤波可以使输入图像中任意方向的细线条保持不变。例如可以选择图 3－15 中的四种线状窗口 $A_1$、$A_2$、$A_3$、$A_4$。用式(3.2.24)的组合式中值滤波可以使输入图像中各种方向的线条保持不变，而且又有一定的噪声平滑性能。



图 3－15　几种线性窗口

(3)其它类型的中值滤波：为了对某些图像，在一定的条件下，尽可能干净地去除噪声，而又尽可能保持有效的图像细节，可以对中值滤波器参数进行某种修正。如迭代中值滤波，就是对输入序列重复进行同样的中值滤波，一直到输出不再有变化为止。如加权中值滤波，也就是

对窗口中的数进行某种加权，以保证滤波的效果。另外中值滤波器还可以和其它滤波器联合使用。总之，图像信息是多种多样的，要求也不一样，因此在处理具体问题时，要依靠丰富的经验来合理有效地使用中值滤波器。

## §3.3 图像的锐化

图像尖锐化处理用于增强图像的边缘及灰度跳变部分，通常讲的勾边增强方法就是图像尖锐化处理。与图像平滑化处理一样，它也有空域和频域两种处理方法。

### 一、微分法

图像模糊的实质就是图像受到平均或积分运算，为实现图像的锐化，必须用它的反运算"微分"，微分运算是求信号的变化率，有加强高频分量的作用，从而使图像轮廓清晰。

为了把图像中间任何方向伸展的边缘和轮廓的模糊变清晰，希望对图像的某种导数运算是各向同性的，可以证明偏导数的平方和运算是各向同性的，梯度和拉普拉斯运算都是符合上述条件的。

1. *梯度法*

对于图像函数 $f(x,y)$，它在点 $f(x,y)$处的梯度是一个矢量，定义为

$$\overline{G}[f(x,y)]=\begin{bmatrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\end{bmatrix} \tag{3.3.1}$$

梯度的两个重要性质是：

(1)梯度的方向在函数 $f(x,y)$最大变化率的方向上。

(2)梯度的幅度用 $G[f(x,y)]$表示，并由下式算出

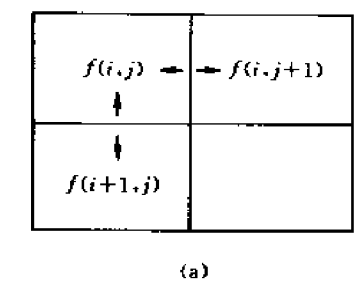
$$G[f(x,y)]=\left[\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2\right]^{1/2} \tag{3.3.2}$$

由上式可知，梯度的数值就是 $f(x,y)$在其最大变化率方向上的单位距离所增加的量。对于数字图像而言，式(3.3.2)可以写为

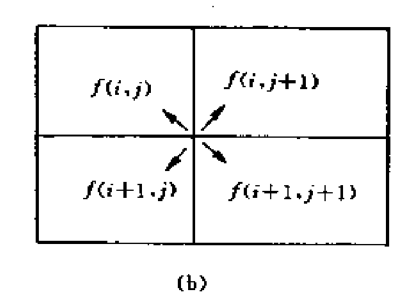
$$G[f(x,y)]=\{[f(i,j)-f(i+1,j)]^2+[f(i,j)-f(i,j+1)]^2\}^{1/2} \tag{3.3.3}$$

式中各象素的位置如图 3-16(a)所示，上式亦可以简化为

$$G[f(x,y)]=|f(i,j)-f(i+1,j)|+|f(i,j)-f(i,j+1)| \tag{3.3.4}$$




(a)




(b)

图 3-16 求梯度的两种差分算法

以上梯度法又叫水平垂直差分法，另一种梯度法叫做罗伯特梯度法(Robert Gradient)，它是一种交叉差分计算法，如图 3-16(b)所示：

$$G[f(x,y)]=\{[f(i,j)-f(i+1,j+1)]^2+[f(i+1,j)-f(i,j+1)]^2\}^{1/2} \quad (3.3.5)$$

同样可近似为：

$$G[f(x,y)]=|f(i,j)-f(i+1,j+1)|+|f(i+1,j)-f(i,j+1)| \quad (3.3.6)$$

以上两种梯度近似算法在图像的最后一行或最后一列的象素的梯度无法标得，一般就用前一行和前一列的梯度值近似代替。

由梯度的计算可知，在图像中灰度变化较大的边沿区域其梯度值大，在灰度变化平缓的区域其梯度值较小，而在灰度均匀区其梯度值为零。图 3-17(a)是一幅二值图像，(b)为计算梯度后的图像经过递度运算后，只留下灰度值急剧变化的边沿处的点。

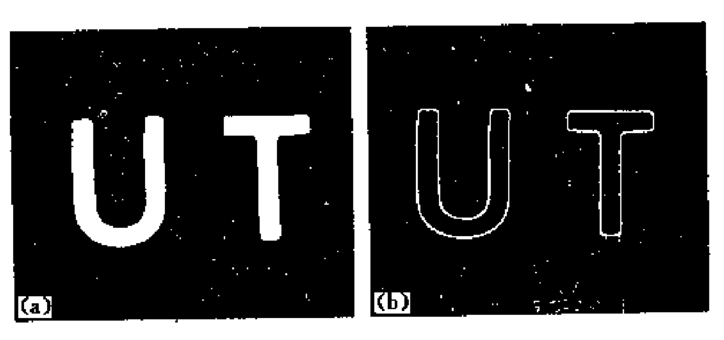



图 3-17

(a)一幅二值图像；(b)计算梯度的结果

当梯度计算完之后，要根据需要生成不同的梯度增强图像。第一种是使各点的灰度 $g(x,y)$ 等于该点的梯度幅度，

$$g(x,y)=G[f(x,y)] \quad (3.3.7)$$

此法的缺点是增强的图像仅显示灰度变化比较徒的边缘轮廓，而灰度变化平缓的区域则呈黑色。

第二种增强的图像是使

$$g(x,y)=\begin{cases}G[f(x,y)] & G[f(x,y)]\geqslant T\\ f(x,y) & \text{其它}\end{cases} \quad (3.3.8)$$

式中 $T$ 是一个非负的阈值，适当选取 $T$，既可使用明显的边缘轮廓得到突出，又不会破坏原来灰度变化比较平缓的背景。

第三种增强图像是使

$$g(x,y)=\begin{cases}L_G & G[f(x,y)]\geqslant T\\ f(x,y) & \text{其它}\end{cases} \quad (3.3.9)$$

式中 $L_G$ 是根据需要指定的一个灰度级，它将明显边缘用一固定的灰度级 $L_G$ 来实现。

第四种增强图像是使

$$g(x,y)=\begin{cases}G[f(x,y)] & G[f(x,y)]\geqslant T\\ L_G & \text{其它}\end{cases} \quad (3.3.10)$$

此法将背景用一个固定灰度级 $L_G$ 来实现，便于研究边缘灰度的变化。

第五种增强图像是使

$$g(x,y)=\begin{cases}L_G & G[f(x,y)]\geqslant T\\ L_B & \text{其它}\end{cases} \quad (3.3.11)$$

此法将背景和边缘用二值图像表示，便于研究边缘所在位置。

2. 拉普拉斯运算

拉普拉斯算子处理是常用的边缘增强处理算子，它是各向同性的二阶导数

$$\nabla^2 f=\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2} \tag{3.3.12}$$

如果图像的模糊是由扩散现象引起的（如胶片颗粒化学扩散，光点散射），则锐化后的图像 $g$ 为

$$g=f-KI\nabla^2 f \tag{3.3.13}$$

式中 $f$、$g$ 分别为锐化前后的图像，$KI$ 为与扩散效应有关的系数。模糊图像 $f$ 经拉普拉斯算子锐化以后得到锐化的不模糊图像 $g$。这里对 $KI$ 的选择要合理，$KI$ 太大会使图像中的轮廓边缘产生过冲。$KI$ 太小、锐化不明显。

对数字图像来讲，$f(x,y)$ 的二阶偏导数表示为：

$$\begin{cases}\dfrac{\partial^2 f(x,y)}{\partial x^2}=\nabla_x f(i+1,j)-\nabla_x f(i,j)\\ \qquad\qquad =[f(i+1,j)-f(i,j)]-[f(i,j)-f(i-1,j)]\\ \qquad\qquad =f(i+1,j)+f(i-1,j)-2f(i,j)\\ \dfrac{\partial^2 f(x,y)}{\partial y^2}=f(i,j+1)+f(i,j-1)-2f(i,j)\end{cases} \tag{3.3.14}$$

为此拉斯算子 $\nabla^2 f$ 为：

$$\begin{aligned}\nabla^2 f &=\frac{\partial^2 f(i,j)}{\partial x^2}+\frac{\partial^2 f(i,j)}{\partial y^2}\\ &=f(i+1,j)+f(i-1,j)+f(i,j+1)+f(i,j-1)-4f(i,j)\\ &=-5\left\{f(i,j)-\frac{1}{5}[f(i+1,j)+f(i-1,j)+f(i,j+1)+f(i,j-1)+f(i,j)]\right\}\end{aligned} \tag{3.3.15}$$

可见数字图像在 $(i,j)$ 点的拉普拉斯算子，可以由 $(i,j)$ 点灰度级值减去该点邻域平均灰度级值来求得。

例 3－4　设有一数字图像 $f(i,j)=1\times n$，其各点的灰度级值为下表所列，计算 $\nabla^2 f$ 及锐化后的各点灰度级值（设 $KI=1$）$g$。

（1）首先按式（3－3－15）计算各点的 $\nabla^2 f$

例如第 3 点：　$\nabla^2 f=-3\left[0-\frac{1}{3}(0+0+1)\right]=1$

　第 8 点：　$\nabla^2 f=-3\left[5-\frac{1}{3}(4+4+5)\right]=-1$

（2）按式（3－3－13）计算 $g=f-\nabla^2 f(KI=1)$

例如第 3 点　$g=f-\nabla^2 f=0-1=-1$

　第 8 点　$g=5-(-1)=6$

将 $f$、$\nabla^2 f$、$g$ 等结果灰度级值，列表如下

| 象素号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $f$ | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 3 |
| $\nabla^2 f$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | −3 | 3 |
| $f-\nabla^2 f$ | 0 | 0 | −1 | 1 | 2 | 3 | 4 | 6 | 5 | 5 | 5 | 5 | 4 | 7 | 6 | 6 | 6 | 6 | 9 | 0 |

（3）从以上例子可以看出，在灰度级斜坡底部和界线的低灰度级侧（如第 3、13、20 点）形成

下冲，在灰度级斜坡顶部和界线的高灰度级侧(8、14、19点)形成上冲。

在灰度平坦区域(如9～12，15～18点)，运算前后没有变化，由此可以看出，拉普拉斯算子可以对由扩散模糊的图像起到边界轮廓增强的效果。要注意，如果不是扩散过程引起的模糊图像，效果并不一定很好。另外，式(3.3.15)又可以表示成模板的形式，如图3－18所示。同梯度算子进行税化一样，拉普拉斯算子也增强了图像的噪声，但跟梯度法相比，拉普拉斯算子对噪声的作用较梯度法弱，故用拉普拉斯算子进行边缘检测时，有必要将图像进行平滑处理。

| 0 | 1 | 0 |
|---|---|---|
| 1 | −4 | 1 |
| 0 | 1 | 0 |

图3－18　拉普拉斯运算模板

## 二、高通滤波

图像中的边缘或线条与图像频谱中的高频分量相对应，因此采用高通滤波器让高频分量顺利通过，使图像的边缘或线条变得清楚，实现图像的锐化。高通滤波可用空域法或频域法来实现。在空间域是用卷积方法，与低通滤波的表达式(3.2.3)一样，只不过其中的冲激响应方阵$H$不同，如以下几种：

$$H=\begin{bmatrix}0&-1&0\\-1&5&-1\\0&-1&0\end{bmatrix}\quad H=\begin{bmatrix}-1&-1&-1\\-1&9&-1\\-1&-1&-1\end{bmatrix}\quad H=\begin{bmatrix}1&-2&1\\-2&5&-2\\1&-2&1\end{bmatrix}$$

高通滤波亦可在频率域中实现，表3－5所示。

表3－5　几种常用的高通滤波传递函数

| 滤波器名称 | $H(u,v)$公式 | $H(u,v)$曲线 |
|---|---|---|
| 理想滤波器 | $H(u,v)=\begin{cases}0 & 如果\ D(u,v)\leqslant D_0\\1 & 如果\ D(u,v)>D_0\end{cases}$ | $H(u,v)$；$D_0$；$D(u,v)$ |
| 巴特沃思滤波器 | $H(u,v)=\dfrac{1}{1+[D_0/D(u,v)]^{2n}}$<br>$H(u,v)=\dfrac{1}{1+[\sqrt{2}-1][D_0/D(u,v)]^{2n}}$ | $H(u,v)$；0.5；1；2；$D(u,v)$ |
| 指数滤波器 | $H(u,v)=\mathrm{e}^{-[D_0/D(u,v)]^n}$<br>$H(u,v)=\exp\left\{-\left[\ln\dfrac{1}{\sqrt{2}}\right]\left[\dfrac{D_0}{D(u,v)}\right]^n\right\}$ | $H(u,v)$；1；0.5；1；2；$D(u,v)$ |
| 梯形滤波器 | $H(u,v)=\begin{cases}0 & 如果\ D(u,v)<D_1\\\dfrac{1}{(D_0-D_1)}[D(u,v)-D] & 如果\ D_1\leqslant D(u,v)\leqslant D_0\\1 & 如果\ D(u,v)>D_0\end{cases}$ | $H(u,v)$；1；$D_1$；$D_0$；$D(u,v)$ |

# §3.4 图像的伪彩色处理

前面讨论的都是单色图像的增强处理，采用伪彩色处理的方法来进行图像增强和显示其效果可以明显的好起来，由于人眼只能区分出由黑到白的十多种到二十多种不同的灰度级，而人眼对彩色的分辨率可以达到几百种甚至上千种。例如大家在观看彩色电视时就可以观察到这一点，在彩色显示时看到了一些画面细节，但当色调一调到黑白显示时，这些细节就会看不出来。所谓伪彩色处理，就是将图像中的黑白灰度级变成不同的彩色，如果分层越多，则彩色越多，人眼所能提取的信息也越多，从而达到图像增强的效果。这是一种视觉效果明显，而又不太复杂的图像增强技术。

伪彩色处理技术不仅适用于航摄和遥感图片，也可以用于 X 光片及云图判读等方面。可以用计算机去做，也可以用专用硬设备来实现。如美国 Digital 电子观察仪 4010，日本 Phosdac－700，1000，1200。国产 NST－1 密度分割伪彩色分析仪，能分出 12 个灰度级以及 12 种彩色显示。一般为了实时观察（如云图判读），多用专用硬件设备。伪彩色处理可以是连续彩色（如彩色电视图），也可以是由几种彩色单独构成。下面我们介绍几种方法的基本原理。

## 一、密度分割——不连续的彩色处理

密度分割是伪彩色处理技术中最简单的一种，设一幅黑白图像 $f(x,y)$，在某一灰度级如 $f(x,y)=l_1$ 上设置一个平行 $XY$ 平面的切割平面。如图 3－19(a)所示黑白图像被切割成只有两个灰度级，对切割平面以下即灰度级小于 $l_1$ 的象素分配给一种颜色（如蓝色），相应地对切割平面上面的即灰度级大于 $l_1$ 的象素分配给另一种颜色（如红色）。这样切割的结果就可以将黑白图像变为只有两个颜色的伪彩色图像，如图 3－19(b)所示。若将以上图像灰度级用 $M$ 个切割平面去切割。就会得到 $M$ 个不同灰度级的区域 $s_1,s_2,\cdots,s_M$。对这 $M$ 个区域中的象素人为地分配以 $M$ 个不同的颜色，就可以得到具有 $M$ 种颜色的伪彩色图像，如图 3－20 所示。密度分割伪彩色处理的优点是简单易行，仅用硬件就可以实现，并可以扩大用途。如计算图像中某灰度级的面积等。但此种方法有些缺点：如伪彩色图像视觉效果不理想，彩色生硬不够调和，且量化噪声大等。为此我们必须研究其它的方法。

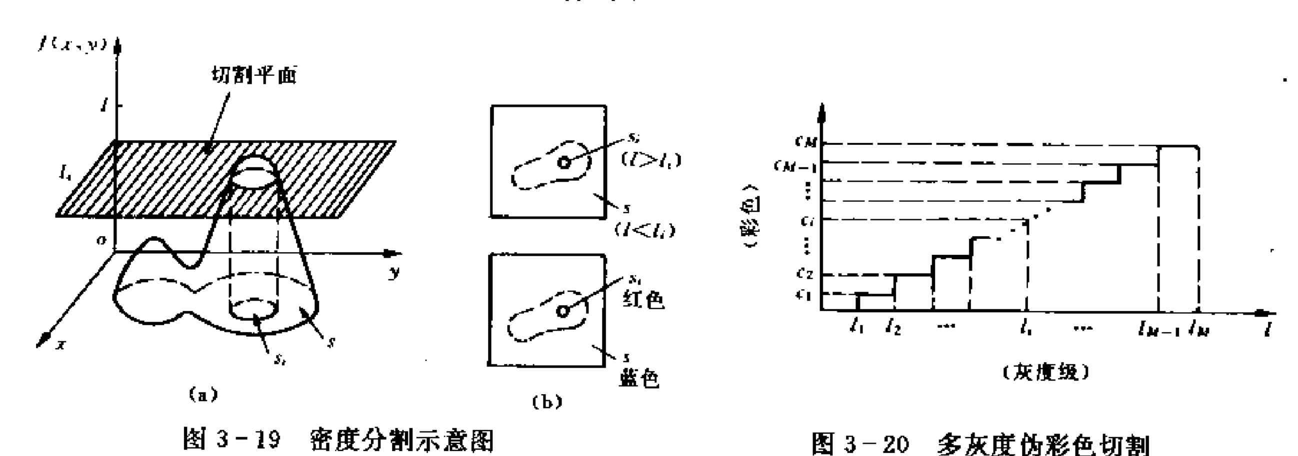



图 3－19 密度分割示意图

图 3－20 多灰度伪彩色切割

## 二、灰度级——彩色变换

这种伪彩色处理技术可以将黑白灰度图像变换为具有多种颜色渐变的连续彩色图像。从实际变换后的图像可以看出其视觉效果明显好于密度分割法。其方法是先将黑白灰度图像送入具有不同变换特征的红、绿、蓝三个变换器。然后再将三个变换器的不同输出分别送到彩色显象管的红、绿、蓝枪。同一灰度由于三个变换器对其实施不同的变换，而使三个变换器输出不同，从而在彩色显象管里合成某一种彩色，可见不同大小灰度级一定可以合成不同彩色。其方框示意图及三个变换器常用变换特性见图 3－21。

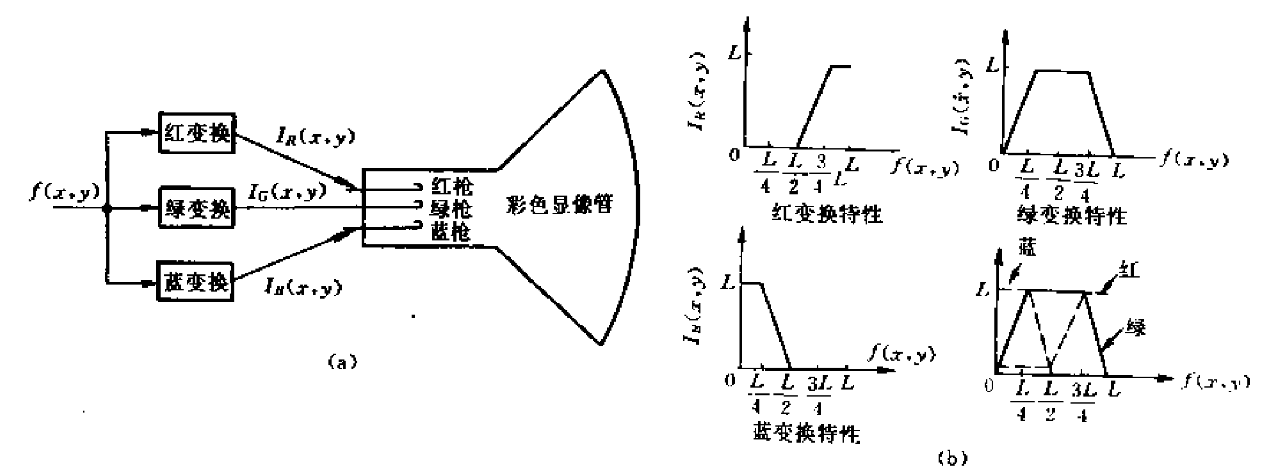



图 3－21　伪彩色变换

从图像中可见，若 $f(x,y)=0$ 则 $I_B(x,y)=L, I_R(x,y)=L_G(x,y)=0$ 从而显示蓝色。同样若 $f(x,y)=\frac{L}{2}$ 则 $L_G(x,y)=L, I_R(x,y)=I_B(x,y)=0$ 从而显示绿色。若 $f(x,y)=L$ 则 $I_R(x,y)=L, I_B(x,y)=I_G(x,y)=0$ 从而显示红色。因此不难理解，若黑白图像 $f(x,y)$ 灰度级在 $0\sim L$ 之间变化，$I_R, I_G, I_B$ 会有不同的输出，从而合成不同的彩色图像。

以上伪彩色处理技术有着广泛的实际应用价值。例如，图像的区域分离显示，彩色印刷制版方面的应用。下面举例说明利用彩色和形状进行微生物分类。

制造大规模集成电路用的纯水，随着大规模集成电路的高密度化，其纯度会要求越来越高。因此不仅要使用多重过滤器除去不纯物中微生物的残骸，而且还要定期对过滤器上残留的微生物进行计数以检查水质。过去，这种工作是人工进行的。现在，为使计数和形状的分类自动化而应用了彩色图像处理技术。

在人工使用显微镜直接检查时，是在通过了一定量水的过滤器(滤纸)上用染色剂对微生物染色来进行的。在明亮的背景中可以得到染有从红到紫各色的，约 1～10 $\mu$m 大小的微生物以及呈黑色微粒的显微镜图像(1500 倍)。过去，人们就用这种方法对每种微生物的数量进行观察和计数。还可以对显微镜图像进行拍照摄影。

用计算机进行计数时，分类的流程如图 3－22 所示。

首先，将 TV 摄象机或扫描鼓的信号数字化，再对输入到计算机中的彩色图像，利用亮度、色度等信号差，从背景中选出微生物区域。这时，利用其色度的不同识别出紫色微生物、红色微生物、黑色微粒子。对已分类的红色微生物根据其面积的不同，除去碎片，剩下来的再根据扁平率的大小区分出球菌、双球菌和杆菌。

## 三、频率滤波法

以上介绍的方法都是在空间域进行的伪彩色处理技术，下面介绍的频率域滤波法，其输出图像的伪彩色与黑白图像的灰度级无关，而仅与黑白图像中的不同空间频率成分有关。它首先是把黑白图像经傅里叶变换到频域，在频率域内用三个不同传递特性的滤波器分离成三个独立分量，再对它们进行反傅里叶变换便得到三幅代表不同频率分量的单色图像，如果希望图像的边缘（高频成分）成为红色，则可以将红色通道滤波器设计成高通滤波器即可，如果要抑制图像中的某种频率成分，则可以把此段频率成分设计成一个带阻滤波器，如

$$H(u,v)=\begin{cases}0 & D(u,v)\leqslant D_0\\ 1 & D(u,v)>D_0\end{cases} \tag{3.4.1}$$

式中 $D(u,v)=\sqrt{(u-u_0)^2+(v-v_0)^2}$，这样在以 $u_0,v_0$ 为中心半径为 $D_0$ 内的所有频率成分被抑制。从图 3－23 可以看出从三个不同频率的滤波器输出的信号再经过傅里叶反变换，可以对这三幅图像再作进一步的处理，如直方图均衡化。最后把它们作为三基色分量分别加到彩色显示器的红、绿、蓝显示通道，从而实现频率域的伪彩色处理，其框图如图 3－23 所示。

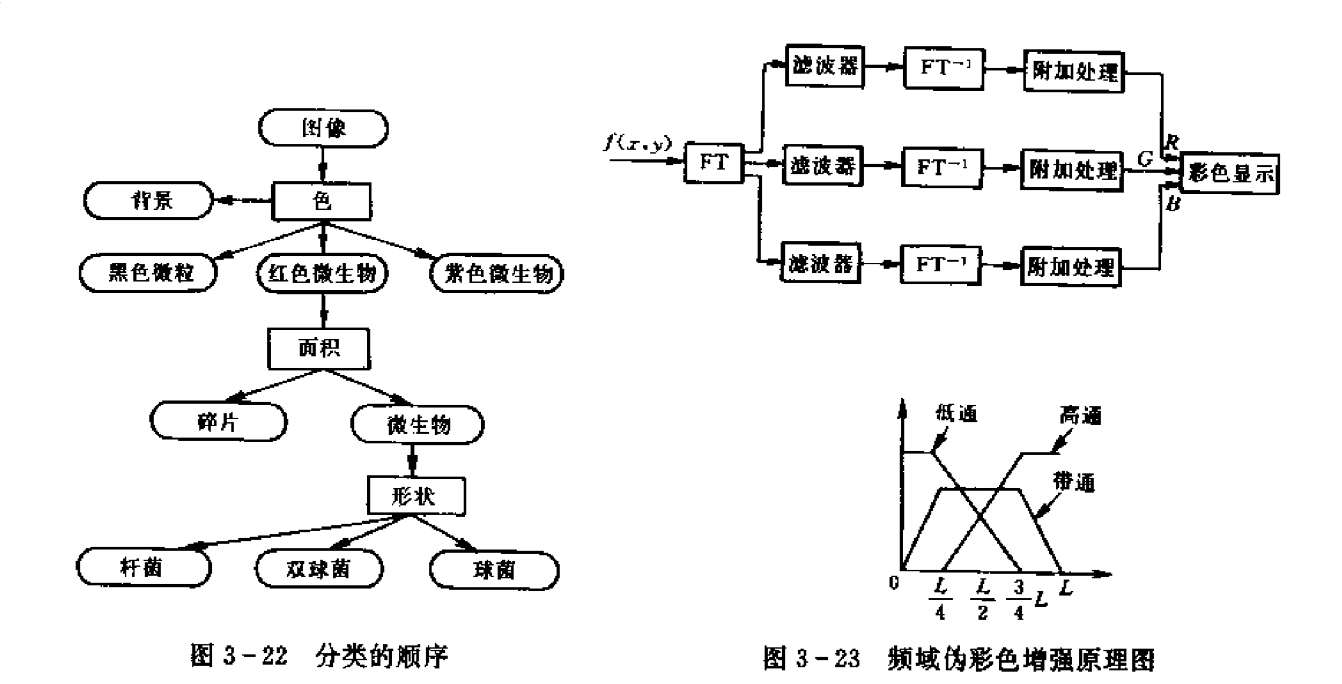



图 3－22　分类的顺序

图 3－23　频域伪彩色增强原理图

# §3.5　图像的几何校正

由于成像系统本身具有非线性或摄象时视角不同，在图像生成的过程中，都会使生成的图像产生几何失真。典型的几种几何失真，如图 3－24 所示。虽然对于卫星遥感图像，其产生几何失真的因素很多很复杂，一般可以分为系统失真和非系统失真。系统失真一般是指由多光谱扫描镜线速不匀，检测器采样延迟造成的各波段间不配准，同一波段扫描行间的错动以及卫星前进运动造成的扫描歪斜等产生的失真，它们一般具有规律性，能预测。非系统失真是指由于卫星飞行姿态的变化（侧滚，俯仰，偏航），飞行高度和速度的变化以及地球自转等引起的失真，如图 3－25 所示。图中虚线为非失真原图，实线为失真图。非系统失真是随机的。

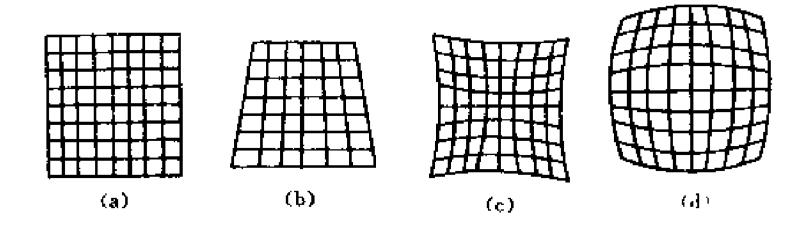
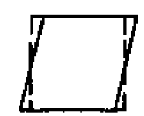


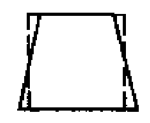

图 3－24　几种典型的几何失真

(a)原图像；(b)透视失真；(c)枕形失真；(d)桶形失真

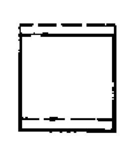
在进行地理制图，土地利用和资源调查时，在许多需要对图像进行定量分析时，首先要对失真的图像先进行精确的几何校正。首先确定一幅图像为基准，然后去校正另一幅图像的几何形状。几何校正一般分两步做：第一步是图像空间坐标的变换；第二步工作必须重新确定在校正空间中各象点的取值。



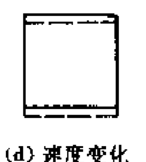
(a) 地球自转



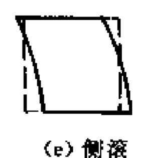
(b) 高度变化



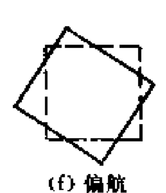
(c) 俯仰



(d) 速度变化



(e) 侧滚



(f) 偏航

图 3－25　非系统几何失真

## 一、图像的空间几何坐标变换

图像的空间几何坐标变换是指按照一幅标准图像 $g(u,v)$ 或一组基准点去校正另一幅几何失真图像 $f(x,y)$。根据两幅图像的一些已知对应点对（又称控制点对），建立起函数关系式，将失真图像的坐标系 $(x,y)$ 变换到标准图像坐标系 $(u,v)$，从而实现失真图像按标准图像的几何位置校正，使 $f(x,y)$ 中的每一象点都可在 $g(u,v)$ 中找到对应象点。通常采用两种方法实现这类坐标变换。

### 1. 三角形线性法

图像的几何失真一般讲是非线性的，但在一个局部小区域内可近似认为是线性的。基于这一假设，将标准图像和被校正图像之间的对应点对划分成一系列小三角形区域，三角形顶点为三个控制点，在三角形区内满足以下线性关系

$$\begin{cases} x = au + bv + c \\ y = du + ev + f \end{cases} \tag{3.4.2}$$

若三对控制点在两个坐标系中的位置分别为 $(x_1,y_1)(x_2,y_2)$、$(x_3,y_3)$ 和 $(u_1,v_1)$、$(u_2,v_2)$、$(u_3,v_3)$，则可建立两级方程组

$$\begin{cases} x_1 = au_1 + bv_1 + c \\ x_2 = au_2 + bv_2 + c \\ x_3 = au_3 + bv_3 + c \end{cases} \tag{3.4.3}$$

$$\begin{cases} y_1 = du_1 + ev_1 + f \\ y_2 = du_2 + ev_2 + f \\ y_3 = du_3 + ev_3 + f \end{cases} \tag{3.4.4}$$

由以上两个方程组可求出 $a,b,c,d,e,f$ 六个系数，再利用式(3.4.2)可实现三角形区内其它象点的坐标变换。对于不同的三角形这六个系数的值是不同的。

这种算法既简单，又能满足一定的精度要求。由于它是以许多小范围内的线性失真去处理

大范围内的非线性失真，所以选择的控制点对越多，分布越均匀，三角形区域的面积越小，则变换的精度越高。当然控制点多又会导致计算量的增加，为此两者之间要折衷考虑。要求控制点尽量覆盖整个待校正的区域，并且控制点的位置要找得准确。

2. 二元多项式法

此方法是将标准图像的空间坐标$(u,v)$和被校正图像的空间坐标$(x,y)$之间的关系用一个二元$n$次多项式来描述

$$x=\sum_{i=1}^{n}\sum_{j=0}^{n-j}a_{ij}u^iv^j \quad y=\sum_{i=0}^{n}\sum_{j=i}^{n-i}b_{ij}u^iv^j \tag{3.4.5}$$

当多项式的次数$n$取为2时：

$$\begin{cases}x=a_{00}+a_{01}v+a_{02}v^2+a_{10}u+a_{11}uv+a_{20}u^2\\ y=b_{00}+b_{01}v+b_{02}v^2+b_{10}u+b_{11}uv+b_{20}u^2\end{cases} \tag{3.4.6}$$

式中$a_{ij}$,$b_{ij}$为待定系数，它可以采用已知的控制点对，用曲面拟合方法，按最小二乘方准则求出。若要使拟合误差平方和$\varepsilon$为最小，即使：

$$\varepsilon=\sum_{e=1}^{L}(x_e-\sum_{i=1}^{n}\sum_{j=0}^{n-i}a_{ij}u_e^iv_e^j)^2=\min$$

则需求：

$$\frac{\partial\varepsilon}{\partial a_{st}}=\sum_{e=1}^{L}2(\sum_{i=1}^{n}\sum_{j=0}^{n-i}a_{ij}u_e^iv_e^j-x_e)u_e^sv_e^t=0 \tag{3.4.7}$$

由此得到：

$$\sum_{e=1}^{L}(\sum_{i=0}^{n}\sum_{j=0}^{n-i}a_{ij}u_e^iv_e^j)u_e^sv_e^t=\sum_{e=1}^{L}x_eu_e^sv_e^t \tag{3.4.8}$$

类似地可以得到

$$\sum_{e=1}^{L}(\sum_{i=0}^{n}\sum_{j=0}^{n-i}b_{ij}u_e^iv_e^j)u_e^sv_e^t=\sum_{e=1}^{L}y_eu_e^sv_e^t \tag{3.4.9}$$

式中$L$为控制点对的个数，$S=0,1,\cdots n,t=0,1\cdots n-s$，以及$s+t\leqslant n$。以上式(3.4.8)与式(3.4.9)为两组由$M$个方程组成的线性方程组，每个方程包含$M$个未知数，$M=(n+1)(n+2)/2$是待求系数的个数。分别解上述二式就可求出$a_{ij}$和$b_{ij}$，代入式(3.4.5)分别解上述二式就可以求出$a_{ij}$和$b_{ij}$，代入式(3.4.5)就可以出现两个坐标系之间的变换。

二元多项式方法比较简单有效，精度较高，精度与所用校正多项式次数有关。多项式次数愈高，位置拟合误差越小。但$n$增加，所需控制点对的数目急剧增加，导致计算时间亦急剧增加。通常用二元二次多项式，控制点对$L\geqslant9$，这时找最小二乘方解。

## 二、象点亮度的确定

图像经几何位置校正后，在校正空间中各象点的亮度值应等于被校正图像对应点的亮度值。但校正后的图像往往会出现某些象点挤压在一起，另一些则分散开了，使校正后的象点不落在坐标点上，为此常采用内插法来求得这些象点的灰度值。常用的方法有三种。

1. 最近邻点法

该法取象点周围四个邻点中距离最近的邻点灰度作为这点的灰度。它计算简单，有一定的精度。缺点是校正后的图像亮度有明显的不连续性。

2. 双线性内插法

本法利用四个邻点的灰度在二个方向上作线性内插。它的算法虽比最近邻法复杂，计算量

大，但结果令人满意，没有灰度不连续性的缺点。双线性内插的示意图如图 3-26 所示。

数学表达式为

$$f(i+u,j+v)=(1-u)(1-v)f(i,j)+(1-u)vf(i,j+1)+ u(1-v)f(i+1,j)+uvf(i+1,j+1) \tag{3.4.10}$$

双线性内插法具有低通滤波性质，它使高频分量受损，图像轮廓模糊。

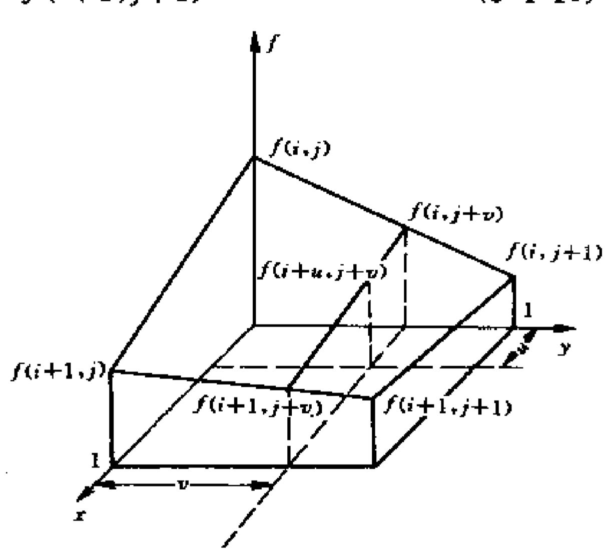



图 3-26 双线性内插法

3. 三次卷积法

此法利用三次多项式 $S(w)$ 来逼近理论上的最佳插值函数 $\mathrm{sinc}(w)$。函数 $S(w)$ 如图 3-27 所示。其数学表达式为

$$\begin{cases} 1-2|w|^2+|w|^3 & |w|<1 \\ 4-8|w|+5|w|^2-|w|^3 & 1\leqslant|w|\leqslant2 \\ 0 & |w|\geqslant2 \end{cases} \tag{3.4.11}$$

计算时利用周围 16 个邻点（图 3-28）的灰度按下式进行内插：

$$f(x,y)=A\cdot B\cdot C$$

$$A=\begin{bmatrix} S(1+v) \\ S(v) \\ S(1-v) \\ S(2-v) \end{bmatrix}^T \qquad C=\begin{bmatrix} S(1+u) \\ S(u) \\ S(1-u) \\ S(2-u) \end{bmatrix}$$

$$B=\begin{bmatrix} f(i-1,j-1) & f(i-1,j) & f(i-1,j+1) & f(i-1,j+2) \\ f(i,j-1) & f(i,j) & f(i,j+1) & f(i,j+2) \\ f(i+1,j-1) & f(i+1,j) & f(i+1,j+1) & f(i+1,j+2) \\ f(i+2,j-1) & f(i+2,j) & f(i+2,j+1) & f(i+2,j+2) \end{bmatrix} \tag{3.4.12}$$

式中的 $S(\cdot)$ 函数用式（3.4.11）计算。此方法计算量大，但能克服前面两种方法的缺点，且精度高。

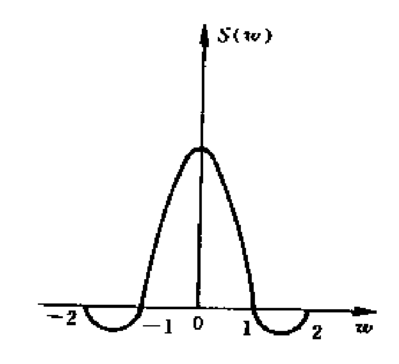



图 3-27 $S(w)$ 函数曲线

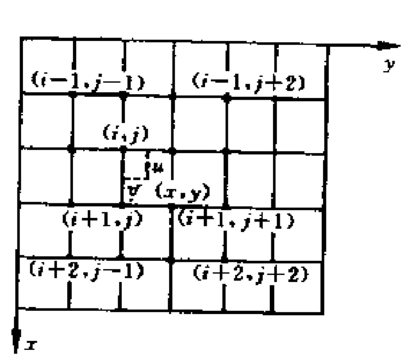



图 3-28 用三次多项式内插

# § 3.6 图像增强实际实现的软硬件框图及结果

## 一、邻域平均法

前面已经谈到了用邻域平均法可以去除图像中的噪声，它是预先设定一个窗口 $A_N$，此窗口可以是方形的，亦可以是其它形状，$N$ 是窗口的大小，即窗口内象素的总数。当此窗口 $A_N$ 沿图像平面平移的过程中，$A_N$ 中心点的灰度值不断地被窗口内的各象素(不包括中心点)的灰度值的平均值所代替，即

$$g(i,j)=\frac{1}{N-1}\sum_{(n+i,m+j)\in A_N} f(n+i,m+j) \tag{3.4.13}$$

当窗口 $A_N$ 从图像的左上角移到图像的右下角时，我们便可以得到平滑后的图像。

一般说来，窗口选择的越大，平滑图像的效果亦越好，但使用邻域平均法是不分青红皂白的一律用周围的点的平均值来替代中心点，这样必然将边缘点也当成噪声点来处理，必然会引起图像边缘的模糊，为此还可以用带门限的邻域平均法及其它方法来进行改进。

用硬件实时实现邻域平均法的总框图如图 3 29 所示。它所选的邻域是 3×3 的矩形窗，其取法如图 3 30 所示。其中心点是$(x,y)$，$A$、取除$(x,y)$点以外的周围八邻点集合。

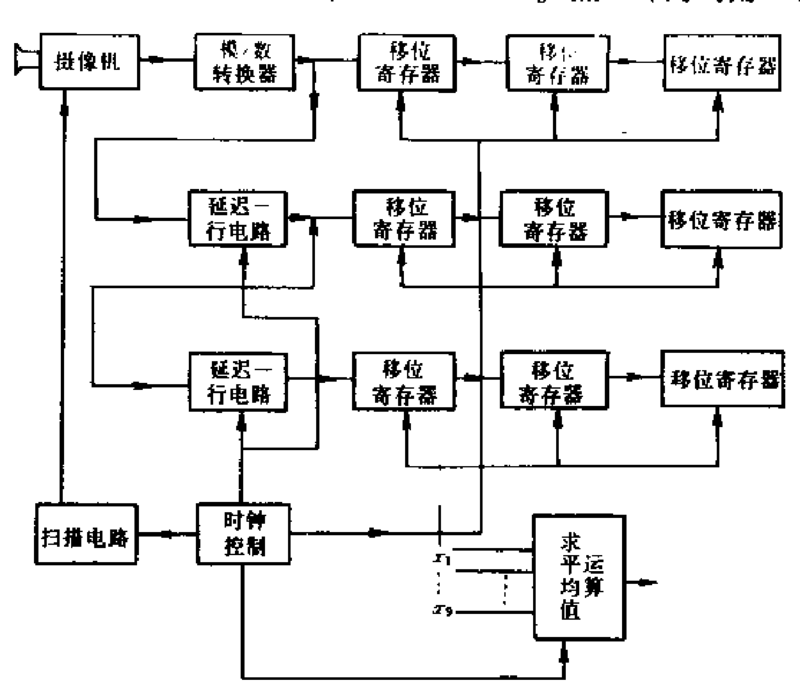



图 3 29 邻域平均法电路框图

现在来看图 3 29，图像信号是由摄象机摄入，经过 A/D 转换器变成了 8 bit 的数字信号，经过寄存器或锁存器延迟一个点，通过延迟一行电路来延迟一行，便能形成图 3 30 的窗。考虑到门延迟时间仅十几个纳秒(ns)，一场信号的周期 $T=20$ ms，处理的图像点是 256×256 阵列，若选用 5 MHz 的间断钟，即处理一个象点的时间是 200 ns，就可以实时的处理从摄象机送入的图像。本电路选用的正是 5 MHz 的间断钟。

延迟一点的电路是利用 74LS374 完成的。

延迟一行电路是 3×3 邻域窗形成的关键所在，在以往的电路结构中，采用的是三个存贮

器结构，其框图见图 3-31。

| $(x-1,y-1)$ | $(x,y-1)$ | $(x+1,y-1)$ |
| --- | --- | --- |
| $(x-1,y)$ | $(x,y)$ | $(x+1,y)$ |
| $(x-1,y+1)$ | $(x,y+1)$ | $(x+1,y+1)$ |

图 3-30　3×3 邻域窗取法

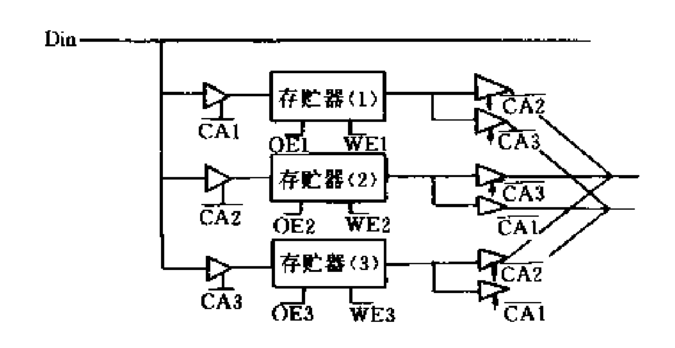



图 3-31　三个存贮器结构形成延迟一行电路

图 3-31 电路的工作原理如下：此电路用了三个存贮器依次存了三行元素，在三个存贮器前加了三个总线驱动器，用写允许信号($\overline{CA1}$,$\overline{CA2}$,$\overline{CA3}$)分别控制其开关，每次只有一个总线开关是打开的，只有在总线开关打开的情况下，它所连接的存贮器才工作，并在写控制信号($\overline{WE1}$,$\overline{WE2}$,$\overline{WE3}$)的控制下进行写操作。与此同时，另两个存贮器同时进行读操作，这种工作状态在三个存贮器之间轮换进行，为实现这一过程，再在每个存贮器的输出端加两个总线驱动器，其开关受读允许信号控制($\overline{OE1}$,$\overline{OE2}$,$\overline{OE3}$)，在读允许的情况下，其存贮器工作，并在读信号($\overline{OE1}$,$\overline{OE2}$,$\overline{OE3}$)控制下进行读操作。读出的二行元素与新的一行元素共同形成三行，经过延迟一点的电路，形成 3×3 的矩形窗。也可以对这种结构进行改进，即仅用二个存贮器便能完成形成三行的任务。在每个延迟一行的电路中仅用一个 256×8 bit 的高速 RAM 即可。其原理图如图 3-32。

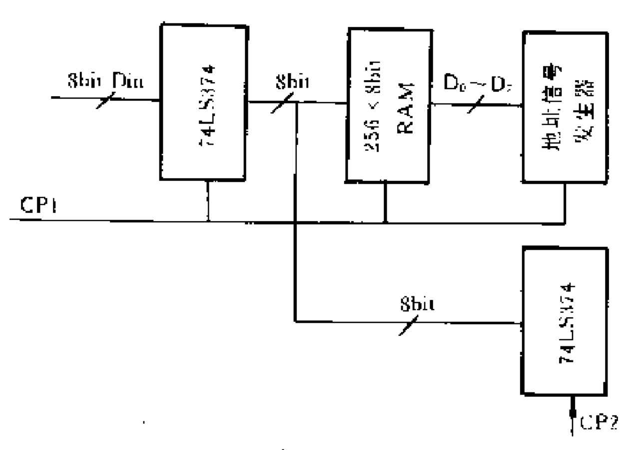



图 3-32　延时一行电路原理图

此电路的工作原理如下，高速的静态 RAM 在一个 200 ns 的周期内将完成一次读操作和一次写操作，其过程是：

(1)在前 100 ns，时钟 $CP_1$ 是高电平，这时存贮器处于读状态，74 LS244 处于高阻状态，在 $CP_1$ 脉冲作用下。存贮器读出了上一行存入高速静态 RAM 的数据，待数据稳定后，在 $CP_2$ 脉冲的作用下，将数据锁存进 74LS374，送下一级使用。

(2)在 $CP_1$ 的后 100 ns，存贮器处于写状态、地址信号未有变动，74LS244 将数据送到数据总线上，在 $CP_1$ 作用下，存贮器将写入新的一行数据，留待下一次作用。在 256 个点过后，静态 RAM 将上一行的数据全部读出，并将现在行写入，这样便完成了延迟一行的任务，两个延迟行与现在行一起，便形成了邻域 3×3 窗所要求的三行。

求均值的运算电路用加法器的级联便能实现，其原理框图见 3—33 图。电路中除 8 运算可用右移三位来实现，由于 74LS283 是 4 bit 的加法器，故用两个搭成 8 bit 的加法器，用三个搭成 12 bit 的加法器。

以上是简单的邻域平均法，由于它无条件运用周围邻点的均值来代替中心点的灰度值，这样就造成了边缘的较大的模糊，根据人眼特性，某象素点的灰度跟背景灰度之差不超过 2% 时，人眼分辨不出这种差别。噪声虽然可能选加在图像上的每一点，但因噪声是均值为零的高

斯噪声，因而大部分的噪声对人眼视觉影响不大，故而没有必要对每一点进行邻域滤波，可用下述算法来处理。

$$g(m,n)\begin{cases} f(m,n) & \text{当} \ |f(m,n)-n| \leqslant T \\ m & \text{当} \ |f(m,n)-n| > T \end{cases} \tag{3.4.14}$$

式中 $m$ 是邻域点的平均值，$T$ 是所加的门限值。

它是先设定一门限 $T$，当中心点灰度跟邻域灰度的均值差大于门限 $T$ 时，认为此点是受噪声污染的点，用邻点的均值来代替该点的灰度，否则认为是图像的细节而加以保留。要想更好的保护图像的边缘，$T$ 选得大一些。为了更好地滤除噪声，$T$ 选得小一些。其软件实现的框图不算困难。

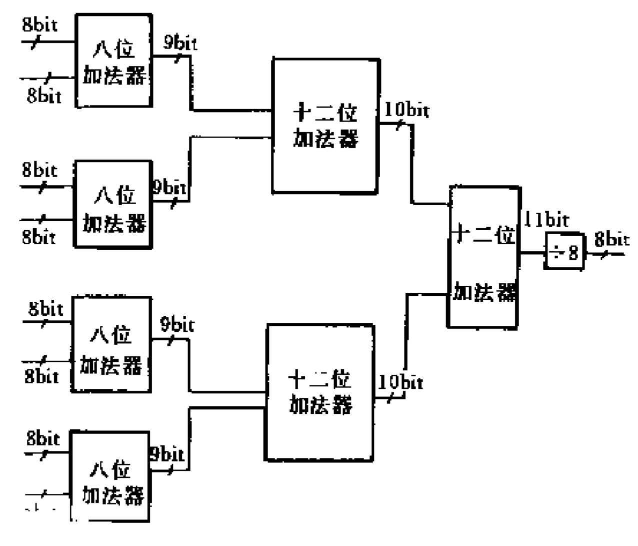


图 3-33　邻域平均的运算电路

## 二、中值滤波算法

由于中值滤波法既能滤除图像的噪声，同时又能保护图像的边缘，而且又便于硬件的实时实现，故而是近年来采用较多的图像平滑算法。

中值滤波的原理前面已经叙述，这里仅叙述结果图像。从对一幅飞机图像处理的结果来看，邻域均值算法的滤波性能最佳，但它对边缘的模糊也很大，用带门限的均值滤波的滤波性能随门限 $T$ 的不同，滤波效果则不同，$T$ 越小，则其滤波效果越好，但对边缘的模糊也较大，$T$ 越大，越能保护边缘，但其滤波性能差。中值滤波的滤除噪声的性能比均值滤波差，但其保护边缘的特点非常突出。

## 三、拉普拉斯算子边缘锐化

前面已叙述了图像锐化是将图像的边缘突出出来，以便于目标的识别和跟踪，我们已经介绍了梯度法(包括水平垂直差分和罗伯特梯度法即交叉差分法)。用以下梯度法进行边缘锐化，可以增强图像的边缘，但是它具有方向性，不能对各种走向的边缘都具有相同的增强效果。而用拉普拉斯算子进行边缘增强。

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial y^2} \tag{3.4.15}$$

用数学方法，完全可以证明 Laplace 算子是各向同性的。拉普拉斯算子的各向同性是区别于其它算法的最大优点之一，即能对任何走向的界线和线条进行锐化，无方向性。

对于数字图像来说，用二阶差分算法来替代拉普拉斯算子，其形式是

$$\begin{aligned}\nabla^2 f &= \nabla_x^2 f + \nabla_y^2 f \\ &= [f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1)] - 4f(i,j)\end{aligned} \tag{3.4.16}$$

亦可以把上式表示成模板的形式如图 3-34，亦可以用一些其它的模板来表示，如图 3-35 所示

下面介绍用图 3－34 的拉普拉斯模板，实现边缘增强的过程，其硬件实现框图见图 3－36。从此图中可以看出，实现此算法的关键仍然是 3×3 矩形邻域窗的形成，其实现框图及其详细说明在邻域平均法中已经介绍。

| 0 | 1 | 0 |
|---|---|---|
| 1 | −4 | 1 |
| 0 | 1 | 0 |

图 3－34　拉普拉斯模板

| 1 | 1 | 1 |
|---|---|---|
| 1 | −8 | 1 |
| 1 | 1 | 1 |

图 3－35　拉普拉斯模板

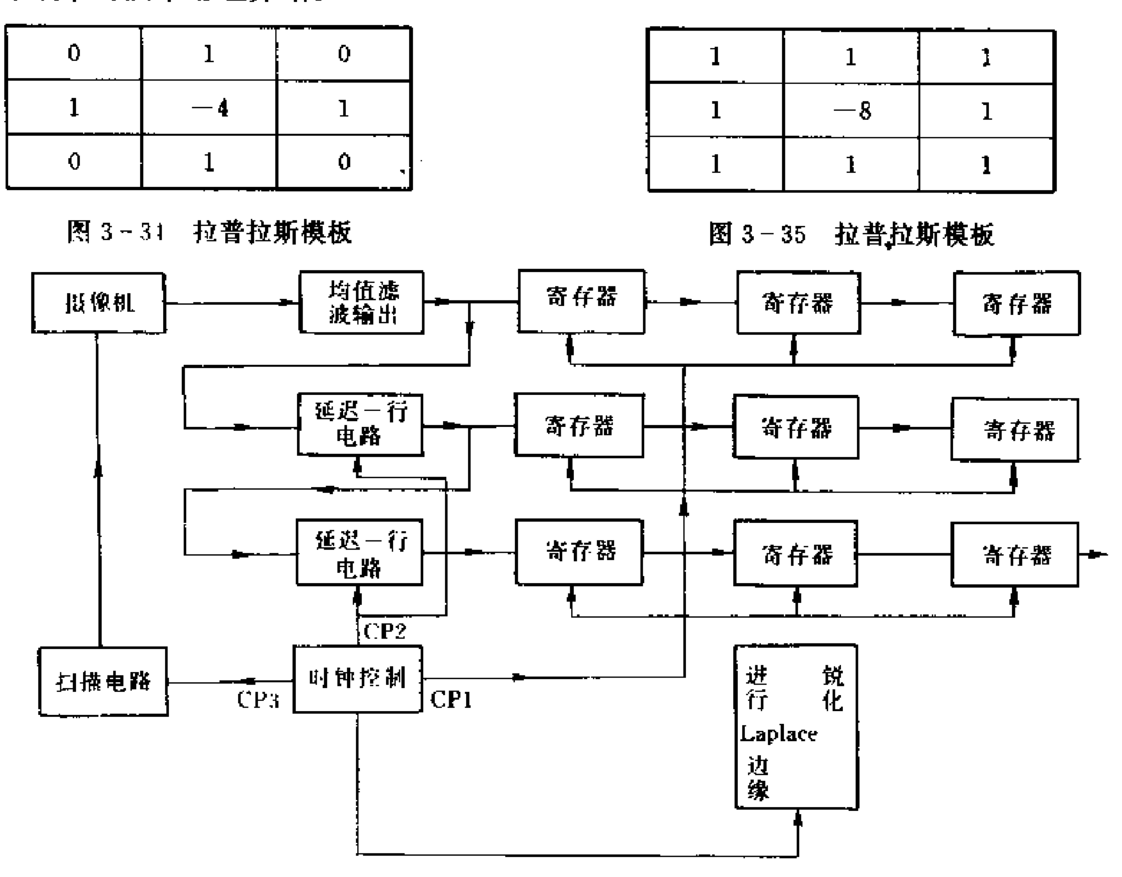



图 3－36　拉普拉斯算子的硬件实现框图

下面讨论拉普拉斯算子的具体计算电路，其实现方案有二。第一种方案见图 3－37。首先从 3×3 窗口中取出 $x_1, x_2, x_3, x_4$ 点，用加法器求其和（记作 $A$），$x_5$ 乘以 4 以后（记作 $B$），$A$ 跟 $B$ 进行比较，并将数据接到 74LS157 和 74LS158 上（由于 157 和 158 均是四位二选一，故用三个搭成 12 位的二选一），用比较器的输出控制 157 和 158 的输出控制端，如 $A>B$，则从 157 输出端输出 $A$，从 158 输出端输出 $\overline{B}$，求 $A+\overline{B}+1$，便得到图像灰度。若 $B>A$，则从 157 输出端输出 $B$，从 158 输出端输出 $\overline{A}$，再求 $B+\overline{A}+1$，便也能求得输出图像的灰度。事实上，此电路求了 $|A-B|$。

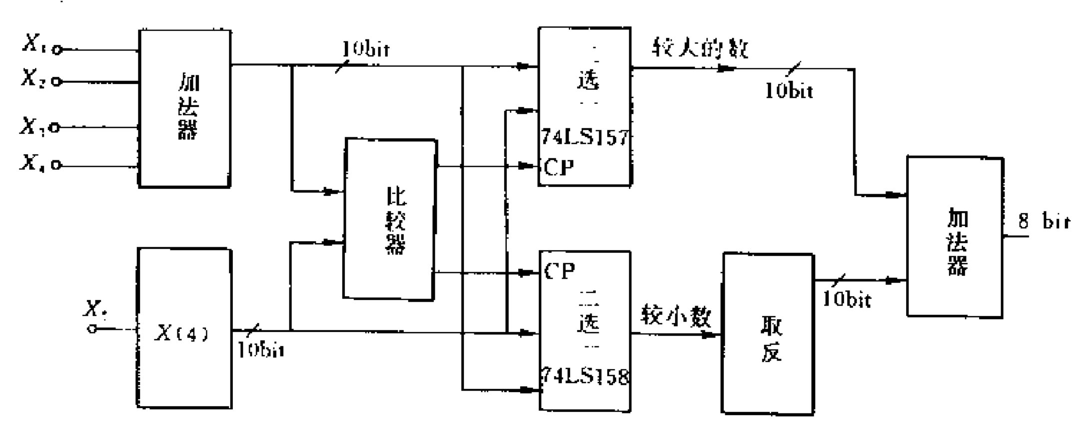



图 3－37　计算拉普拉斯算子的电路框图

第二种方案，见图 3－38。首先从 3×3 的邻域窗中取出 $x_1$，$x_2$，$x_3$，$x_4$ 和 $x_5$ 点，$x_1$，$x_2$，$x_3$，$x_4$ 经过加法器求和，其值记作 $A$，$x_5$ 乘以 4 以后通过数据线送到比较器和算术逻辑单元，其和记作 $B$。由于采用的算术逻辑单元有 $A-B$ 和 $B-A$ 的功能，故 $A$ 跟 $B$ 进行比较，用比较器输出控制算术逻辑运算单元，若 $A$ 大于 $B$，则运算逻辑单元输出 $A-B$ 的值，若 $B$ 大于 $A$ 的话，则算术运算单元输出 $B-A$ 的值，这样图 3－38 的电路便完成了求 $|A-B|$ 的功能。显然用方案二实现要比方案一简便得多。

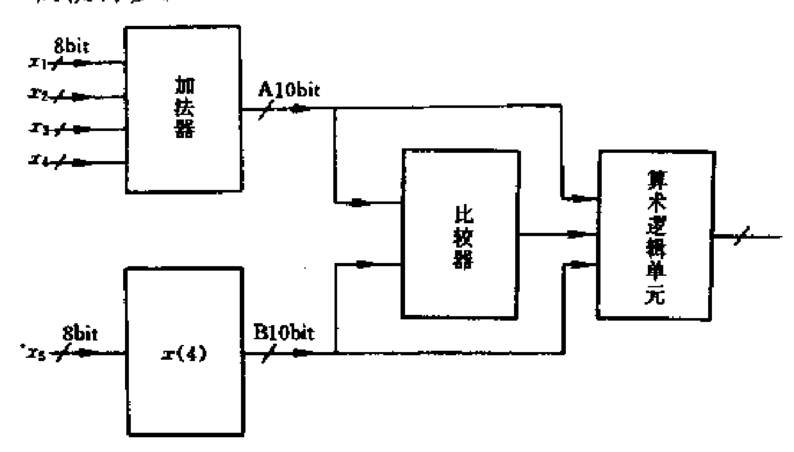



图 3－38　拉普拉斯算子计算电路

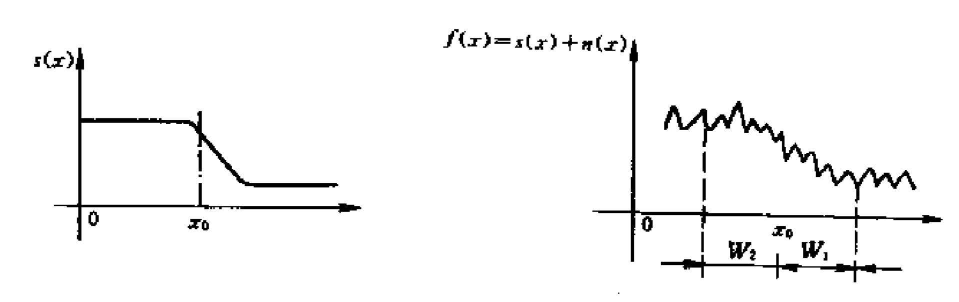



图 3－39　一个理想的和实际的边缘波形

除了梯度算子、拉普拉斯算子以外，再简单介绍一个 sobel 算子。梯度锐化将会使噪声，条纹等得到增强，sobel 算子则在一定程度上克服了这个问题。以一维的情况来说明，对于一个理想的边缘，实际情况总是迭加有噪声，如图 3－39 所示。要求出实际波形 $f(x)$ 的边缘，一种比较合理的方法是求出 $W_2$ 窗内的平均（见图 3－39），再求出 $W_1$ 窗内元素的平均，然后再对这个平均值求梯度，窗 $W_1$ 和 $W_2$ 的选择应该不能将图像 $s(x)$ 的跳变完全模糊掉，即定义如下算法。

$$D(x)=\frac{1}{W_1}\int_x^{x+W_1} f(u)\mathrm{d}u-\frac{1}{W_2}\int_x^{x+W_2} f(u)\mathrm{d}u$$

| | | |
|---|---|---|
| $a$ | $b$ | $c$ |
| $d$ | $f(i,j)$ | $f$ |
| $g$ | $h$ | $i$ |

图 3－40

显然这种方法将平均和差分结合起来，同样若是将这种方法推广到二维的情形，便得到了 Sobel 算子法。Sobel 算子法的算法原理是，假设有一个 3×3 的矩阵，如图 3－40 所示，将按下述算法变换图像的灰度，变换后图像的灰度是：

$$g=\sqrt{S_x^2+S_y^2} \tag{3.4.17}$$

式中　$S_x=(c+2f+i)-(a+2d+g)$

$S_y=(g+2h+i)-(a+2b+c)$

为了简化计算，用 $g=|S_x|+|S_y|$ 来代替式(3.4.17)的计算，从而得到锐化后的波形。

从上面的讨论可知，Sobel 算子不象普通梯度算子那样用两个象素之差值，而是用两列或两行加权和之差值，这就导致了两个优点：

(1)由于引入了平均因素，因而对图像中的随机噪声有一定的平滑作用。

(2)由于它是相隔两行或两列之差分，故边缘两侧之元素得到了增强，故边缘显得粗而且亮。

下面列出对原始图像进行各种处理的实验结果如图 3-41 到图 3-45 所示。从结果中可以看出，几种方法均能勾勒出图像的边缘，但在不同的程度上增加了图像的噪声，其中 Sobel 算子对噪声的增强作用小，具有一定的平滑特性。用拉普拉斯算子和梯度算子对噪声的增强作用较大。

从各向同性观点来看，拉普拉斯算子是典型的各向同性算子，对各方向的增强强度是一样的，梯度算子和 Sobel 算子都具有一定的方向性。

从边缘增强的效果来看，用 Sobel 算子增强的图像边缘粗而亮，这主要是由于 Sobel 算子采用了两行或两列加权和之差值。用拉普拉斯算子勾勒边缘较梯度算子强，但较 Sobel 算子弱。



图 3-41　原始图像

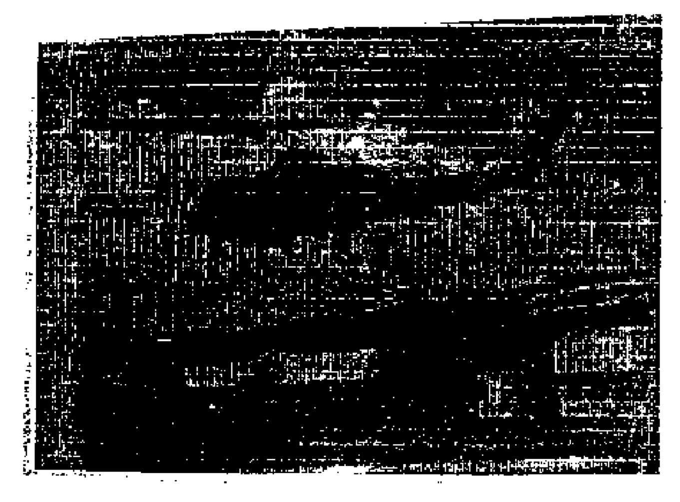


图 3-42　加门限的均值滤波　此处限 $T=15$

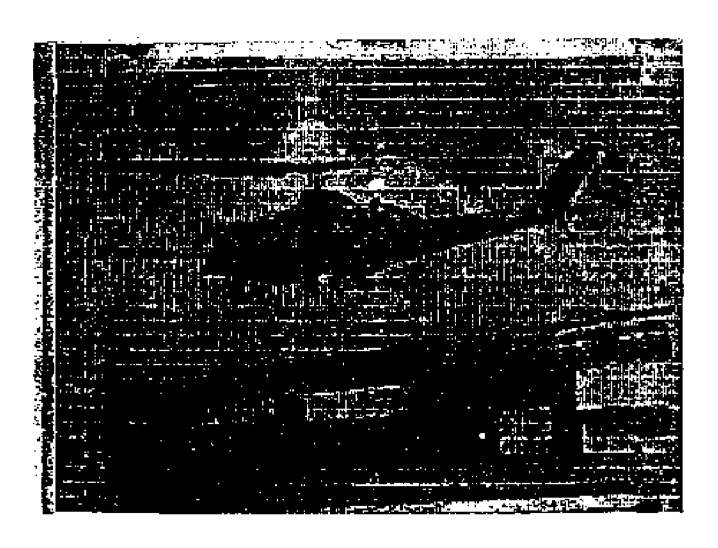


图 3-43　中值滤波算法



图 3-44　用 Sobel 算子进行边缘增强

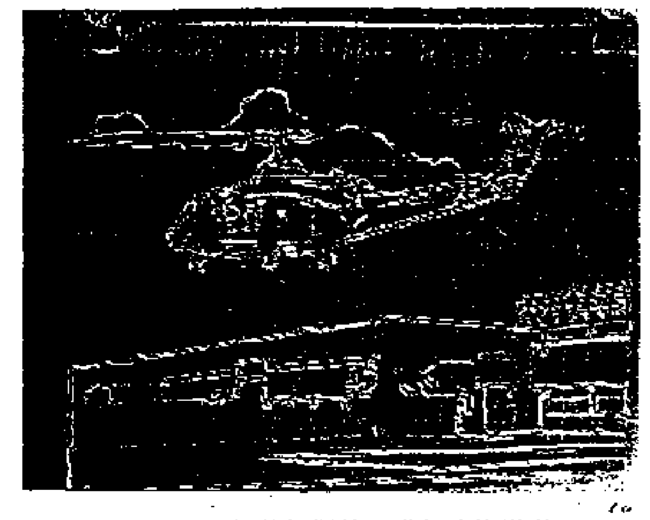


图 3-45　用拉普拉斯算子进行边缘增强

## 习　　题

3-1　试给出把灰度范围(0,10)拉伸为(0,15),把灰度范围(10,20)移到(15,25),并把灰

度范围(20,30)压缩为(25,30)的变换方程。

3—2　试给出变换方程$t(Z)$,使其满足在$10\leqslant Z\leqslant 100$的范围内,$t(Z)$是$\lg Z$的线性函数。

3—3　已知一幅64×64,3 bit数字图像,各个灰度级出现的频数如表3-1(a)所示。要求将此幅图像进行直方图变换,使其变换后的图像具有表3-1(b)所示的灰度级分布。并画出变换前后图像的直方图,以资比较。

表3-1(a)

| $f(x,y)$ | $n_k$ | $n_k/n$ |
|---|---|---|
| 0 | 560 | 0.14 |
| 1 | 920 | 0.22 |
| 2 | 1046 | 0.26 |
| 3 | 705 | 0.17 |
| 4 | 356 | 0.09 |
| 5 | 267 | 0.06 |
| 6 | 170 | 0.04 |
| 7 | 72 | 0.02 |

表3-1(b)

| $g_k(x,y)$ | $n_k$ | $n_k/n$ |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 790 | 0.19 |
| 4 | 1023 | 0.25 |
| 5 | 850 | 0.21 |
| 6 | 985 | 0.24 |
| 7 | 448 | 0.11 |

3—4　已知一幅图像如图3-46所示,即半边为深灰色,其灰度级为1/7,而另半边是黑色,其灰度级为0,假定(0,1)之间划分为8个灰度等级,试对此图像进行均匀化处理,并描述一下均匀化后的图像是一幅什么样的图像。

图3-46　题3—4图

3—5　有一幅图像如图3-47所示,由于干扰,在接收时图中有若干个亮点(灰度为255),试问此类图像如何处理?并将处理后的图像画出来。

| | | | | | |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 8 | 7 | 4 |
| 2 | 255 | 2 | 3 | 3 | 3 |
| 3 | 3 | 255 | 4 | 3 | 3 |
| 3 | 3 | 3 | 255 | 4 | 6 |
| 3 | 3 | 4 | 5 | 255 | 8 |
| 2 | 3 | 4 | 6 | 7 | 8 |

图3-47

3—6　试证明拉普拉斯算子$\left(\frac{\partial^2}{\partial x^2}\right)+\left(\frac{\partial^2}{\partial y^2}\right)$的旋转不变性。

3—7　图像增强中数字拉普拉斯算子常用什么形式?试用拉普拉斯算子对图3-48进行增强运算,并将增强后的图像画出来。

| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

图 3-48 题 3—7 图

# 参 考 文 献


[3—1] Iterative Historam Modification IEEE Trans. SMC—8 April 1978;IEEE Trans,SMC —8 July 1978.

[3—2] Gold B. and Rader C. M. 信号的数字处理。地质出版社,1980

[3—3] D. CC Wang,Vagnucci AH,Licc,Gradient inverse Weighted Somthing Scheme and evaluation of its Performancs CGIP 6:167-187

[3—4] 陈廷标等编著. 数字图像处理. 北京:人民邮电出版社,1990

# 第四章　图像的复原

在成像过程中，由于成像系统各种因素的影响，可能使图像质量降低。这种图像质量的降低，被称之为“退化”。与图像增强相似，图像复原的目的也是改善图像的质量。但是图像复原是试图利用退化过程的先验知识使已被退化的图像恢复本来面目，而图像增强是用某种试探的方式改善图像质量，以适应人眼的视觉和心理，这是图像复原与图像增强的本质差别。因此，本章讨论的基础是退化的数学模型，图像复原可以看成是退化的逆过程。

在成像系统中，引起图像退化的原因很多。例如，成像系统的散焦，成像设备与物体的相对运动，成像器材的固有缺陷以及外部干扰等。由这些因素造成的图像退化的典型现象是图像模糊，去模糊是一种基本的复原问题。由于图像复原是建立在比较严格的数学推导之上，所以有较多而复杂的数学运算是本章的特点。

## §4.1　退化的数学模型

设有一成像系统，当输入为 $f(x,y)$ 时，输出为 $g(x,y)$，即

$$g(x,y)=T[f(x,y)] \tag{4.1.1}$$

运算 $T[\,]$ 表示这一系统对输入图像的作用，它可能是线性的，也可能是非线性的。现在讨论线性系统的情况。

一幅图像 $f(x,y)$ 可以看作是由一系列点源组成的，因此，$f(x,y)$ 可以通过点源函数的卷积来表示。

$$f(x,y)=\iint_{-\infty}^{\infty} f(\alpha,\beta)\delta(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{4.1.2}$$

式中 $\delta$ 函数为点源函数，且有

$$\delta(x-\alpha,y-\beta)=\delta(x-\alpha)\delta(y-\beta) \tag{4.1.3}$$

因此，经成像系统处理后，输出图像为

$$\begin{aligned} g(x,y)=T[f(x,y)]&=\iint_{-\infty}^{\infty} f(\alpha,\beta)T[\delta(x-\alpha,y-\beta)]\mathrm{d}\alpha\mathrm{d}\beta \\ &=\iint_{-\infty}^{\infty} f(\alpha,\beta)h(x,\alpha,y,\beta)\mathrm{d}\alpha\mathrm{d}\beta \end{aligned} \tag{4.1.4}$$

式中 $h(x,\alpha,y,\beta)=T[\delta(x-\alpha,y-\beta)]$ 称为系统的点扩散函数，或系统的冲击响应。图像退化正是由于系统的冲击响应是位移不变的，这时

$$h(x,\alpha,y,\beta)=h(x-\alpha,y-\beta) \tag{4.1.5}$$

因此输出图像为

$$g(x,y)=\iint_{-\infty}^{\infty} f(\alpha,\beta)h[(x-\alpha,y-\beta)]\mathrm{d}\alpha\mathrm{d}\beta=f(x,y)*h(x,y) \tag{4.1.6}$$

如果$T[\,]$是可分离系统，则

$$h(x,\alpha,y,\beta)=h_1(x,\alpha)h_2(y,\beta) \tag{4.1.7}$$

一个系统如果满足这个条件，就可以把二维问题化作一维问题，这与正交变换中可分离核是相似的。

对一个线性、位移不变、可分离系统，它所成的像可表示为

$$g(x,y)=\iint_{-\infty}^{\infty} f(\alpha,\beta)h_1(x-\alpha)h_2(y-\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{4.1.8}$$

图像复原就是对某一系统而言，已知$g(x,y)$，求$f(x,y)$的问题。

图像退化除了成像系统本身的因素以外，有时还受到噪声的污染。一般假定噪声$n(x,y)$是加性白噪声。这时退化后的图像为

$$g(x,y)=\iint_{-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta+n(x,y) \tag{4.1.9}$$

图像退化的示意图如图4－1所示。



图4－1　图像的退化模型

## §4.2　连续系统的图像复原

### 一、一般原理

式(4.1.9)所示退化模型可以简写成

$$g=T[f]+n$$

图像复原就是求上式的解$f$，如果有解，则可以写成

$$f=T^{-1}[g-n] \tag{4.2.1}$$

对于一个特定的图像来说，其解应当是唯一的。但是可以证明，如果不附加条件，式(4.1.9)的解并不是唯一的。因为我们总可以找到一个函数$l(x,y)$和$h(x-\alpha,y-\beta)$正交，即

$$\iint_{-\infty}^{\infty} h(x-\alpha,y-\beta)l(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta=0 \tag{4.2.2}$$

因此，$f(x,y)+c\,l(x,y)$都是这个积分方程的解，所以解不是唯一的。

另外，在研究傅里叶级数时，有这样一个定理，如果$\Psi(x)$在任意区间内可积，而且

$$\int_{-\infty}^{\infty} |\Psi(x)| \mathrm{d}x < \infty$$

则

$$\lim_{\lambda \to \infty} \int_{-\infty}^{\infty} \Psi(x) \sin(\lambda x) \mathrm{d}x = 0 \tag{4.2.3}$$

$$\lim_{\lambda \to \infty} \int_{-\infty}^{\infty} \Psi(x) \cos(\lambda x) \mathrm{d}x = 0$$

所以只要 $\iint_{-\infty}^{\infty} |h(x-\alpha, y-\beta)| \mathrm{d}\alpha \mathrm{d}\beta$ 对任意的 $x, y$ 有界，则对于任意小的数 $\varepsilon$，总能找到一个数 $A$，当 $\lambda_1, \lambda_2 > A$ 时，使积分

$$\iint_{-\infty}^{\infty} h(x-\alpha, y-\beta) \sin(\lambda_1 \alpha) \sin(\lambda_2 \beta) \mathrm{d}\alpha \mathrm{d}\beta < \varepsilon \tag{4.2.4}$$

因此式(4.1.9)这样的积分方程是没有唯一解的，只能在许多解中找出满足特定条件的合理解。根据不同的特定条件可以形成不同的复原方法，下面将分别进行讨论。

## 二、逆滤波

设成像系统是线性时不变的，则

$$g(x,y) = \iint_{-\infty}^{\infty} f(\alpha, \beta) h(x-\alpha, y-\beta) \mathrm{d}\alpha \mathrm{d}\beta + n(x,y) \tag{4.2.5}$$

对式(4.2.5)两边作傅里叶变换得

$$G(u,v) = F(u,v) H(u,v) + N(u,v) \tag{4.2.6}$$

式中 $G(u,v)$、$F(u,v)$、$H(u,v)$、$N(u,v)$ 分别是 $g(x,y)$、$f(x,y)$、$h(x,y)$、$n(x,y)$ 的傅里叶变换。$H(u,v)$ 又称为系统的转移函数。

在没有噪声的理想情况下，式(4.2.6)变为

$$G(u,v) = F(u,v) H(u,v) \tag{4.2.7}$$

在上式两边乘以 $1/H(u,v)$ 得

$$F(u,v) = \frac{G(u,v)}{H(u,v)} \tag{4.2.8}$$

因此已知系统的转移函数 $H(u,v)$，用上式即可求得 $F(u,v)$，由此可以通过傅里叶反变换得到原始函数

$$f(x,y) = F^{-1}\left[\frac{G(u,v)}{H(u,v)}\right] \tag{4.2.9}$$

这种复原方法就称为逆滤波。频域的逆滤波模型如图 4－2 所示。

$F(u,v)$ ⟶ [ $H(u,v)$ ] ⟶ [ $1/H(u,v)$ ] ⟶ $F(u,v)$

图 4－2　频域的逆滤波模型(无噪声)

但是，在实际情况下总有噪声存在，若仍用逆滤波方法，则从式(4.2.6)可得 $F(u,v)$ 的估

值如下式

$$\hat{F}(u,v)=F(u,v)+\frac{N(u,v)}{H(u,v)} \tag{4.2.10}$$

逆滤波复原的原理十分简单，但是由于在实现中存在较大的困难而使复原效果受到很大影响。首先是$H(u,v)$零点的影响。在无噪声的理想情况下，如果$H(u,v)$在某些对图像信号有较大影响的点或区域上为零，那么$G(u,v)$的值在这些频率处也为零，因此就无法用式(4.2.8)确定这些频率处$F(u,v)$的值。而存在噪声的一般情况下，从式(4.2.10)可见，估值$\hat{F}(u,v)$和实际值$F(u,v)$之间相差一项$N(u,v)/H(u,v)$。显然在$H(u,v)=0$的频率上，由于噪声的存在，使这些频率上的估值$\hat{F}(u,v)$没有意义。而且在这些零点附近，由于$H(u,v)$非常小，$N(u,v)/H(u,v)$变得很大，从而误差大大增加，以致无法得到正确的复原结果。

在一般情况下，$H(u,v)$具有低通特性，即在原点附近的有限区域内$H(u,v)\neq 0$，而且迅速下降。而噪声则由于有较宽的带宽，下降的速度要慢得多，因此逆滤波复原常在原点附近的有限区域内进行，以避免$H(u,v)$出现零点或有较小的值。这样可以得到比较合理的结果，但是仍然不可避免地要对复原后的图像质量产生较大的影响。

### 三、维纳滤波

我们仍然考虑线性时不变系统，并重写式(4.2.5)如下

$$g(x,y)=\iint_{-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha,y-\beta)\mathrm{d}\alpha\mathrm{d}\beta+n(x,y) \tag{4.2.11}$$

现在求这一方程的解$f(x,y)$的估值$\hat{f}(x,y)$。设它有如下形式

$$\hat{f}(x,y)=\iint_{-\infty}^{\infty} m(x-\alpha,y-\beta)g(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{4.2.12}$$

假设$g(x,y)$(因此也包括$f(x,y)$和$n(x,y)$)是平稳的随机变量，$h(x,y)$是确知的函数。现在求$m(x,y)$，使误差

$$e^2=E[|f(x,y)-\hat{f}(x,y)|^2] \tag{4.2.13}$$

最小。这就是著名的维纳滤波问题，解法很多，这里直接给出连续的维纳滤波的最后结果。

设$S_{ff^*}(u,v)$为$f(x,y)$的相关函数$R_{ff^*}(x,y)$的傅里叶变换。$S_{gg^*}(u,v)$是$g(x,y)$的相关函数$R_{gg^*}(x,y)$的傅里叶变换。$S_{nn^*}(u,v)$是噪声$n(x,y)$的相关函数$R_{nn^*}(x,y)$的傅里叶变换。$H(u,v)$是点扩展函数$h(x,y)$的傅里叶变换，则有

$$\begin{aligned} M(u,v)&=\frac{H^*(u,v)S_{f^*f}(u,v)}{S_{g^*g}}=\frac{H^*(u,v)S_{f^*f}(u,v)}{H(u,v)H^*(u,v)S_{f^*f}(u,v)+S_{n^*n}(u,v)} \\ &=\frac{|H(u,v)|^2}{H(u,v)[|H(u,v)|^2+S_{n^*n}(u,v)/S_{f^*f}(u,v)]} \end{aligned} \tag{4.2.14}$$

式中 $S_{f^*f}(u,v)=S_{ff^*}(-u,-v)$；$S_{g^*g}(u,v)=S_{gg^*}(-u,-v)$

$$S_{n^*n}(u,v)=S_{nn^*}(-u,-v)$$

这就是维纳滤波的解，式中$S_{n^*n}(u,v)/S_{f^*f}(u,v)$是噪声和信号能量之比。若噪声$n(x,y)=0$，则上式变为$M(u,v)=1/H(u,v)$，即为逆滤波。为了计算$S_{n^*n}$，可以令输入$f(x,y)=0$，则$g$

$(x,y)=n(x,y)$，由此计算 $R_{n^*n}(x,y)$，并得 $S_{n^*n}=F[R_{n^*n}(x,y)]$。当解 $M(u,v)$满足式4.2.14时，均方误差为

$$e^2 = \iint_{-\infty}^{\infty} \frac{S_{n^*n}(u,v)}{|H(u,v)|^2 + S_{n^*n}(u,v)/S_{f^*f}(u,v)} \mathrm{d}u\mathrm{d}v \tag{4.2.15}$$

## §4.3　离散情况下的退化模型

前一节讨论了在连续情况下，对一个线性时不变系统而言，退化模型可以用连续函数的卷积来表示。同样，在离散情况下，退化模型可以表示成离散卷积的形式。下面先讨论一维的情况。

### 一、一维

为使讨论简化，不考虑噪声的存在，这时退化模型可用下式离散卷积来表示。设离散序列 $f(m)$定义在 $m=0,1,\cdots,N-1$ 各点上，在其它 $m$ 值上 $f(m)=0$，$h(m)$定义在 $m=0,1,\cdots,M-1$ 各点之上，在其它 $m$ 值上，$h(m)=0$；且 $N>M$，则它们的离散卷积为

$$g(m) = \sum_{k=0}^{N-1} f(k)h(m-k) = f(m) * h(m) \tag{4.3.1}$$

由此得到结果 $g(m)$在 $m=0,1,\cdots,M+N-2$ 上有确定的值，在其它 $m$ 值上 $g(m)=0$。上式还可以用矩阵的形式表示。

离散卷积还可用周期卷积来表示。首先将 $f(m)$和 $h(m)$分别作周期延拓，周期都为 $P=M+N-1$，即在一个周期内

$$f_e(m)=\begin{cases} f(m) & 0\leqslant m\leqslant N-1 \\ 0 & N\leqslant m\leqslant P-1 \end{cases} \tag{4.3.2}$$

$$h_e(m)=\begin{cases} h(m) & 0\leqslant m\leqslant M-1 \\ 0 & M\leqslant m\leqslant P-1 \end{cases} \tag{4.3.3}$$

周期延拓是为了避免卷积的各个周期产生交迭。则周期卷积可定义为：

$$f_e(m) \otimes h_e(m) = \sum_{k=0}^{P-1} f_e(k)h_e((m-k))_P \tag{4.3.4}$$

式中 $m$ 为任意值，$((k))_P=k \bmod p$，在一个周期内

$$f_e(m)\otimes h_e(m)=f(m) * h(m)=g(m)$$

即

$$g(m) = \sum_{k=0}^{P-1} f_e(k)h_e(m-k) \qquad (m=0,1,\cdots,P-1) \tag{4.3.5}$$

用矩阵表示，则为

$$\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(P-1) \end{bmatrix} = \begin{bmatrix} h_e(0) & h_e(-1)\cdots h_e(-P+1) \\ h_e(1) & \cdots \quad \cdots h_e(-P+2) \\ \vdots & \vdots \quad \vdots \quad \vdots \\ h_e(P-1) & h_e(P-2) \cdots \quad h_e(0) \end{bmatrix} \begin{bmatrix} f_e(0) \\ f_e(1) \\ \vdots \\ f_e(P-1) \end{bmatrix} \tag{4.3.6}$$

因为 $h_e(m)$的周期为 $P$,所以 $h_e(m+P)=h_e(m)$,因此

$$\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(P-1) \end{bmatrix} = \begin{bmatrix} h_e(0) & h_e(P-1)\cdots h_e(1) \\ h_e(1) & \cdots \quad \cdots h_e(2) \\ \vdots & \vdots \quad \vdots \quad \vdots \\ h_e(P-1) & \vdots \quad \cdots h_e(0) \end{bmatrix} \begin{bmatrix} f_e(0) \\ f_e(1) \\ \vdots \\ f_e(P-1) \end{bmatrix} \tag{4.3.7}$$

还可以写成更简洁的形式

$$g=[h]f \tag{4.3.8}$$

式中 $g$ 和 $f$ 都是 $P$ 维列向量,$[h]$则是 $P\times P$ 阶矩阵。由式(4.3.7)可见,矩阵$[h]$中每一行元素均相同,只是每行以循环方式右移一位,因此矩阵$[h]$称为循环矩阵。可以证明,循环矩阵相加还是循环矩阵;循环矩阵相乘还是循环矩阵。

### 二、二维

设一离散二维信号 $f(m,n)$,它在 $M\times N$ 大小的范围内有值,其它位置上为零,即

$$f(m,n)=\begin{cases} f(m,n) & 0\leqslant m\leqslant M-1 \quad 且\ 0\leqslant n\leqslant N-1 \\ 0 & 其它 \end{cases}$$

且设

$$h(m,n)=\begin{cases} h(m,n) & 0\leqslant m\leqslant J-1 \quad 且\ 0\leqslant n\leqslant n-1 \\ 0 & 其它 \end{cases}$$

这时二维卷积可以写成

$$g(m,n)=\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}h(m-k,n-l)f(k,l)=h(m,n)*f(m,n) \tag{4.3.9}$$

结果 $g(m,n)$在 $P\times Q$ 范围内有值,$P=M+J-1$,$Q=N+K-1$,即有

$$g(m,n)=\begin{cases} g(m,n) & 0\leqslant m\leqslant P-1 \quad 且\ 0\leqslant n\leqslant Q-1 \\ 0 & 其它 \end{cases} \tag{4.3.10}$$

这说明卷积结果不为零的区域相当于在 $f$ 的两个边上加上宽度为 $J-l$ 和 $K-l$ 的区域。

和一维情况一样,可以用二维周期卷积代表二维的线性卷积,这时同样要对 $f(m,n)$和 $h(m,n)$作周期延拓,延拓后的大小为 $P\times Q$,同样有

$$P=M+J-1, \qquad Q=N+K-1$$

即

$$f_e(m,n)=\begin{cases} f(m,n) & 0\leqslant m\leqslant M-1 \quad 且 \quad 0\leqslant n\leqslant N-1 \\ 0 & M\leqslant m\leqslant P-1 \quad 或 \quad N\leqslant n\leqslant Q-1 \end{cases} \tag{4.3.11}$$

$$h_e(m,n)=\begin{cases} h(m,n) & 0\leqslant m\leqslant J-1 \quad 且 \quad 0\leqslant n\leqslant K-1 \\ 0 & J\leqslant m\leqslant P-1 \quad 或 \quad K\leqslant n\leqslant Q-1 \end{cases}$$

则 $h_e(m,n)$ 和 $f_e(m,n)$ 的周期卷积为

$$\begin{aligned} g_e(m,n) &= \sum_{k=0}^{P-1}\sum_{l=0}^{Q-1} h_e[((m-k))_P,((n-l))_Q]f_e(k,l) \\ &= h_e(m,n)\otimes f_e(m,n) \end{aligned} \tag{4.3.12}$$

式中 $m,n$ 为任意值。在一个周期内，即当 $0\leqslant m\leqslant P-1$ 和 $0\leqslant n\leqslant Q-1$，周期卷积的结果与线性卷积相同，即

$$g_e(m,n)=h(m,n)*f(m,n)=g(m,n) \tag{4.3.13}$$

现在用向量 $g$ 和 $f$ 来表示二维离散图像 $g(m,n)$ 和 $f(m,n)$，方法是将 $g(m,n)$ 和 $f(m,n)$ 中的元素按行堆砌成列向量：

$$g=[g(0,0),g(0,1),\cdots,g(0,Q-1),g(1,0),\cdots,g(1,Q-1),\cdots,g(P-1,Q-1)]^T \tag{4.3.14}$$

用同样的方法堆砌成 $f$，这样就可以把式(4.3.12)写成矩阵形式

$$g=[h]f \tag{4.3.15}$$

式中 $[h]$ 是 $PQ\times PQ$ 矩阵，可以写成

$$[h]=\begin{bmatrix} [h_0] & \cdots & [h_1] \\ [h_1] & & [h_2] \\ \vdots & & \vdots \\ [h_{P-1}] & & [h_0] \end{bmatrix} \tag{4.3.16}$$

式中 $[h_j]$ 为子矩阵，大小为 $Q\times Q$，即 $[h]$ 矩阵由 $P\times P$ 个大小为 $Q\times Q$ 的子矩阵组成。

子矩阵的下标是以循环的方式排列的，其中

$$[h_j]=\begin{bmatrix} h_e(j,0) & h_e(j,Q-1) & \cdots & h_e(j,1) \\ h_e(j,1) & h_e(j,0) & \cdots & h_e(j,2) \\ \vdots & \vdots & \cdots & \vdots \\ h_e(j,Q-1) & h_e(j,Q-2) & \cdots & h_e(j,0) \end{bmatrix} \tag{4.3.17}$$

显然，矩阵 $[h_j]$ 中元素的第二个下标也是以循环方式变化的，所以 $[h_j]$ 也是循环矩阵。因此，矩阵 $[h]$ 被称为分块循环矩阵。

有关二维卷积的运算，这里不再多叙，望参考其它书籍。

## §4.4 离散情况下的复原

本节叙述的离散情况下的复原方法是建立在前面讨论的退化模型的基础上，它们的基本思想都是设法找出原始图像的估计值 $\hat{f}$，使预先确定的某个优化准则最小。这些方法可分为有约束条件复原和无约束条件复原。

### 一、无约束条件复原

这里仅讨论逆滤波复原的情况。

前面已经证明，成像系统的输出图像和输入图像的关系可以用堆砌以后的向量方程来表示

$$g=[h]f+n \tag{4.4.1}$$

此式比式(4.3.8)多了一项噪声项 $n$，这是考虑到实际情况而加上的。式中$[h]$为一分块循环矩阵。现在求一个估值向量 $\hat{f}$，使 $g-[h]\hat{f}$ 的幅值最小，或看作使噪声最小。为此定准则函数为

$$\begin{aligned}J[\hat{f}]&=[g-[h]\hat{f}]^T[g-[h]\hat{f}]\\&=g^Tg-\hat{f}^T[h]^Tg-g^T[h]\hat{f}+\hat{f}^T[h]^T[h]\hat{f}\end{aligned} \tag{4.4.2}$$

为了求得 $\hat{f}$，使准则函数最小，可将上式对向量 $\hat{f}$ 求导，并令其等于 0。下面用到对向量求导的两个性质，设 $a$ 和 $b$ 两个向量，$[A]$为对称矩阵，$J$ 为一标量，则

(1) 如果 $J=a^Tb=b^Ta$，则

$$\frac{\partial J}{\partial a}=b \tag{4.4.3}$$

(2) 如果 $J=a^T[A]a$，则

$$\partial J/\partial a=2[A]a \tag{4.4.4}$$

现在对式(4.4.2)求导，并令其等于零，则有

$$\frac{\partial J}{\partial \hat{f}}=-[h]^Tg-[g^T[h]]^T+2[h]^T[h]\hat{f}=0$$

即

$$-[h]^Tg+[h]^T[h]\hat{f}=0 \tag{4.4.5}$$

若$[h]$的逆$[h]^{-1}$存在，则

$$\{[h]^T[h]\}^{-1}=[h]^{-1}\{[h]^T\}^{-1}$$

代入式(4.4.5)得

$$\hat{f}=[h]^{-1}\{[h]^T\}^{-1}[h]^Tg=[h]^{-1}g$$

或

$$g=[h]\hat{f}$$

对上式作傅里叶变换，得

$$G(u,v)=H(u,v)\cdot \hat{F}(u,v) \tag{4.4.6}$$

由此得离散情况下的逆滤波公式

$$\hat{F}(u,v)=G(u,v)/H(u,v) \tag{4.4.7}$$

这是不考虑噪声情况下的逆滤波公式，它与连续情况下的结果有相同的形式。

### 二、有约束条件复原

维纳滤波是将图像看作是一个平稳随机过程，以图像和噪声的相关矩阵为基础，在统计意义上求最小误差而得到的。因此它的结果在平均意义上最佳。现在讨论在确定意义上是最佳的，即把 $f$ 看作一个确知解。为了求得合理解，需要加上一个约束条件，设 $Q$ 为 $f$ 的线性算子。即，使

$$\|[Q]f\|^2=\{[Q]f\}^T[Q]^f=f^T[Q]^T[Q]^f \tag{4.4.8}$$

最小，同时满足图像的噪声能量为常数的约束条件，即

$$\|g-[h]\hat{f}\|^2=\{g-[h]\hat{f}\}^T\{g-[h]\hat{f}\}=C$$

为了得到符合上述要求的解，利用拉格朗日乘子法得准则函数为

$$J[\hat{f}]=\|[Q]\hat{f}\|^2+\alpha[\|g-[h]\hat{f}\|^2-C]$$

式中，$\alpha$、$C$ 为常数。同样将上式对 $\hat{f}$ 求导，并令其为零，就可得到要求的解。

$$\frac{\partial J}{\partial \hat{f}}=2[Q]^T[Q]\hat{f}-2\alpha[h]^T\{g-[h]\hat{f}\}=0$$

由此解得

$$f=\{[h]^T[h]+\frac{1}{\alpha}[Q]^T[Q]\}^{-1}[h]^Tg \tag{4.4.9}$$

与逆滤波一样，为简化计算，通常把它转到频率域去，经过复杂的计算，由于篇幅有限，仅把结果列出如下：

$$\hat{F}(u,v)=\frac{H^*(u,v)G(u,v)}{|H(u,v)|^2+\frac{1}{\alpha}|Q(u,v)|^2} \tag{4.4.10}$$

式中 $\alpha$ 是一个未定参数，$\alpha$ 与噪声有关。设 $\beta=1/\alpha$，$\beta$ 为待定的参量。为了满足约束条件 $\|g-[h]\hat{f}\|^2=\|n\|^2$，应调整参量 $\beta$ 直至这一条件得到满足。

以上有约束复原方法又称为最小二乘方滤波复原。

## §4.5 运动模糊图像的复原

在照相的曝光期间，如果物体和相机有相对运动，则反映在底片上的图像有明显的移动，形成了模糊的运动图像。

设运动方向为 $X$ 轴方向则图像函数可以写成 $f(x-x_0(t),y)$，这里 $x_0(t)$是沿 $x$ 的正方向运动的距离，设曝光时间为 $0\leqslant t\leqslant T$，在此期间移动的距离是 $a$，所以 $x_0(t)=at/T$。形成的模糊图像为

$$g(x,y)=\alpha\int_0^T f(x-\frac{at}{T},y)\mathrm{d}t \tag{4.5.1}$$

$\alpha$ 是与照片感光灵敏度有关的系数，现已知 $g(x,y)$ 和 $\alpha$，求 $f(x,y)$。由于模糊只有在 $x$ 方向，而和 $y$ 方向无关。所以可以一行一行地复原。在某一行 $y=y_0$ 时，把 $g(x,y_0)$ 和 $f(x,y_0)$，简写成 $g(x)$ 和 $f(x)$，则有

$$g(x)=\alpha\int_0^T f(x-\frac{at}{T})\mathrm{d}t \qquad (4.5.2)$$

作变量代换，令 $x'=x-\frac{at}{T}$，并对 $g(t)$ 求导得(过程略)

$$f(x)=\beta g'(x)+f(x-a) \qquad (4.5.3)$$

式中 $\beta=a/T\alpha$，上式说明 $f(x)$ 有递推性质。



图 4-3　图像的运动模糊

模糊图像 $g(x)$ 的导数是可以求得的，所以已知 $f(x-a)$ 就可以推得 $f(x)$。实际只要求出总长度为 $a$ 的区间上的图像值，就可以推得整个图像了。

设研究区域为 $0\leqslant x\leqslant L$，$L=Ka$，$K$ 为正整数，即将这一区域分成 $K$ 段长度为 $a$ 的子区域，$m$ 是段的数目，通过推演可以得到，移动图像的一般表示式为

$$f(Z+ma)=\beta\sum_{k=0}^{m}g'(Z+ka)+\Psi(Z) \qquad 0\leqslant Z\leqslant a \qquad (4.5.4)$$

式中不论 $m$ 为何值，$\Psi(Z)$ 总是存在的，因此它是一个周期函数，当 $ma\leqslant x\leqslant(m+1)a$ 时

$$f(x)=\beta\sum_{k=0}^{m}g'(x-Ka)+\Psi(x-ma) \qquad 0\leqslant x\leqslant L \qquad (4.5.5)$$

上式中 $g'$ 可以由模糊图像求得，但 $\Psi(Z)$ 是一未知函数，对于任何一个 $\Psi(Z)$，都有一个满足式(4.5.1)的解，所以只有确定了某一长度为 $a$ 的区间的图像以后才能确定 $\Psi(Z)$，下面介绍一种粗糙的近似方法。

由式(4.5.4)有

$$\Psi(Z)=f(Z+ma)-\beta\sum_{k=0}^{m}g'(Z+ka)$$

式中 $m=0,1,\cdots,K-1$ 共有 $K$ 项，将上式以 $K$ 项相加得

$$K\Psi(Z)=\sum_{m=0}^{K-1}f(Z+ma)-\beta\sum_{m=0}^{K-1}\sum_{k=0}^{m}g'(Z+ka)$$

$$\Psi(Z)=\frac{1}{K}\sum_{m=0}^{K-1}f(Z+ma)-\frac{\beta}{K}\sum_{m=0}^{K-1}\sum_{k=0}^{m}g'(Z+ka) \qquad (4.5.6)$$

上式中第一项当 $K$ 值较大时，趋向于 $f$ 的平均值，因此可将它近似地看作一个常数，所以设

$$\frac{1}{K}\sum_{m=0}^{K-1}f(Z+ma)=A$$

则式(4.5.6)可以写成

$$\begin{aligned}\Psi(Z)&=A-\frac{\beta}{K}\sum_{m=0}^{K-1}\sum_{k=0}^{m}g'(Z+ka)\\&=A-\frac{\beta}{K}\{Kg(Z)+(K-1)g'(Z+a)+\cdots+g'[Z+(k-1)a]\}\end{aligned} \qquad (4.5.7)$$

这里 $A$ 是一个未知数，要用试验方法确定，将上式代入式(4.5.4)得

$$f(Z+ma)\approx A+\beta\sum_{k=0}^{m}g'(Z+ka)-\frac{\beta}{K}\sum_{m=0}^{K-1}\sum_{k=0}^{m}g'(Z+ka)$$

最后可得

$$f(x)\approx A+\beta\sum_{k=0}^{m}g'(x-ka)-\frac{\beta}{K}\sum_{m=0}^{K-1}\sum_{k=0}^{m}g'(x-ka) \qquad (4.5.8)$$

以上讨论的运动模糊图像的复原只考虑了 $x$ 方向的匀速直线运动，但利用上述原理不难得到一个同时考虑两个方向上匀速直线运动的去模糊表达式。

下面将讨论把运动模糊与前面的复原方法联系起来，以得到另一类运动模糊图像的复原方法。

图像退化的数学模型可以写为

$$g(x,y)=h(x,y)*f(x,y)+n(x,y)$$

它们的傅里叶变换表示为 $G(u,v)=H(u,v)F(u,v)+N(u,v)$

其中 $n(x,y)$ 和 $N(u,v)$ 是噪声项，与运动无关，现将图像按二个方向运动，$a_i+b_i=v$，所以可以将运动模糊图像表示成积分。

$$g(x,y)=a\int_0^T f(x-at,y-bt)\mathrm{d}t+n(x,y) \tag{4.5.9}$$

上式中

$$\begin{aligned}&\int_0^T f(x-at,y-bt)\mathrm{d}t\\&=\int_0^T\left[\iint_{-\infty}^{\infty}F(u,v)\exp\{\mathrm{j}2\pi[u(x-at)+v(y-bt)]\}\mathrm{d}u\mathrm{d}v\right]\mathrm{d}t\\&=\iint_{-\infty}^{\infty}F(u,v)\frac{\mathrm{e}^{\mathrm{j}\pi(au+\beta v)}\sin[\pi(au+bv)T]}{\pi(au+bv)}\mathrm{e}^{\mathrm{j}2\pi(ux+vy)}\mathrm{d}u\mathrm{d}v\end{aligned}$$

式(4.5.9)的频域表示为

$$G(u,v)=F(u,v)H(u,v)+N(u,v)$$

式中 $$H(u,v)=\alpha\mathrm{e}^{-\mathrm{j}\pi(\alpha u+bv)}\sin[\pi(au+\beta v)T]/\pi(\alpha u+bv) \tag{4.5.10}$$

有了 $H(u,v)$，就可以用维纳或逆滤波等方法来复原匀速直线运动形成的模糊图像，由式(4.5.10)可以看出，$H(u,v)$ 是一种常见的特殊函数 $\sin x/x$ 再加上移项组成的，从这一特点可以看出它包含着运动方向和距离的信息。

# §4.6 非线性图像复原

## 一、最大后验恢复

维纳滤波在恢复中引入了统计方法，并使用了均方差判据。另一种统计方法是把原图像 $f(x,y)$ 和退化图像 $g(x,y)$ 都作为随机场，在已知 $g(x,y)$ 的前提下，求出后验条件概率密度函数 $P(f(x,y)/g(x,y))$。设 $\hat{f}$ 是当 $P(f/g)$ 取最大值时的 $f$，则 $\hat{f}(x,y)$ 就代表已知退化图像 $g(x,y)$ 时最可能的原图像 $f(x,y)$，也即它是 $f$ 的最大后验估计。

当考虑到图像的非线性退化时，退化模型可表示为

$$g(x,y)=S[b(x,y)]+n(x,y) \tag{4.6.1}$$

$$b(x,y)=\iint_{-\infty}^{\infty}h(x,\alpha,y,\beta)]f(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{4.6.2}$$

由贝叶斯准则知道 $P(f/g)P(g)=P(g/f)P(f)$。因此求 $P(f/g)$ 的最大值等效于求

$$\max_f \frac{P(g/f)P(f)}{P(g)} \approx \max_f P(g/f)P(f) \tag{4.6.3}$$

对以上贝叶斯公式(4.6.3)可以做如下的比喻，假如 $P(g)$ 表示发生肝炎病的可能性，$P(f)$ 为由于某种原因引起肝炎病的可能性，则肝病发生后由某种原因引起肝炎的条件概率 $P(f/g)$ 与 $P(f)$ 和 $P(g/f)$ 的乘积有关。$P(f)$ 是某因素，例如输血、环境和饮食等引起肝炎发生的先验概率，$P(g/f)$ 是某因素会引发肝炎的条件概率，$P(g)$ 一般为常数。在图像恢复中 $P(f/g)$ 代表含噪图像中确认是信息图像 $f$ 的可能性，要进行恢复需要有关 $f$ 的统计模型，通过适当选择 $f$ 求出最大化的 $P(f/g)$ 来完成恢复，这就是最大后验恢复法的过程。

最大后验法把图像看作是非平稳随机场，可以把图像模型表示成一个平稳随机过程对于一个不平稳的均值作零均值 Gaussian 起伏。这种 Gaussion 模型可导出一个简单的方程组来近似表示式(4.6.3)求最大值问题。最大值问题的解可通过迭代法求出最佳值。有多种迭代式可用，有建议使用下面的迭代序列。

$$\hat{f}_{k+1} = \hat{f} - h * S_b\{\sigma_n^{-2}[g - S(h * \hat{f}_k)]\} - \sigma_f^{-2}(\hat{f}_k - \bar{f}) \tag{4.6.4}$$

这里 $k$ 为迭代次数，$S_b$ 为由非线性函数 $S$ 的导数组成的函数，$\sigma_f^{-2}$ 和 $\sigma_n^{-2}$ 分别为 $f$ 和 $n$ 的方差的倒数，$\bar{f}$ 是随空间而变的均值，经验表明它是一个常数，但要经过多次迭代才收敛到最后的解。

式(4.6.4)表明一个恢复图像可以通过一个序列的卷积(式中用符号 $*$ )来估算，因此任何卷积空间滤波的程序都可用来估算式(4.6.4)，此式也可以当作一个非线性维纳滤波通过迭代收敛到最后的解。即使 $S$ 为线性时，有时使用式(4.6.4)也是有益的，由于我们通过序列卷积得到恢复的图像，我们可以在完全收敛之前选择一个合适的解。

有关最大后验估算法，还可以用广泛使用的组合优化算法、模拟退火法来实现。

根据贝叶斯估计定理得到的后验概率公式为

$$P(X=x/Y=y) = \frac{P(Y=y/X=x) \times P(X=x)}{P(Y=y)} \tag{4.6.5}$$

其中

$$P(X=x) = \frac{1}{Z}\exp\{-\frac{1}{T}\sum_{c \in C} V_C(x)\}$$

它代表信息场 $X$ 的联合概率分布。式中 $Z$ 是配分函数，是一个归一化的常数；$T$ 是温度，为一控制参数；$C$ 为所求象点周围邻点的组合；$V_C(x)$ 为所有邻域点的势函数。含噪图像的模型 $P(Y=y/X=x)$ 可以写为

$$P(Y=y/X=x) = \prod_{(i,j)\in L} \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y_{ij} - x_{ij})^2\} \tag{4.6.7}$$

$\sigma^2$ 为噪声方差。将式(4.6.7)和式(4.6.6)代入式(4.6.5)可以得到

$$P(X=x/Y=y) = \frac{1}{Z_p}\exp[-E(x)] \tag{4.6.8}$$

$Z_p$ 为规一化常数，$E(x)$ 为能量函数

$$E(x) = -\frac{1}{T}\sum_{c \in C} V_c(x) + \sum_{(ij)\in L} \frac{1}{2\sigma^2}(y_{ij} - x_{ij})^2 = E_1(x) + E_2(x) \tag{4.6.9}$$

上式中 $E_1(x)$ 受图像本身子集的影响，而 $E_2(x)$ 受高斯噪声的影响，对一个信噪比固定的含噪图像而言，$E_1(x)$ 是固定不变的，影响后验概率最大的主要因素就是 $E_2(x)$，因此图像恢复的过

程就是寻找系统能量最小的过程。

模拟退火法主要是与物理中固体物质的退火过程类同，当系统的温度 $T$ 缓慢降低时，其内能逐步地降低。当温度达到足够低时，系统将达到内能最小值，在图像恢复的过程中，当式(4.6.8)中的系统内能达到最低值时，则系统的条件概率 $P(X=x/Y=y)$ 将达到最大值，此刻即为我们所求的图像恢复的最大值。以上模拟退火进行图像恢复的执行过程如下。

首先给出一个初始状态 $\{x\}$，计算它的能量 $E[\{x\}]$，让系统随机地变化到状态 $\{x_j\}$，计算 $\Delta E$，以如下的概率接受新组态。

$$P=\begin{cases}1 & \text{如果 } \Delta E<0 \\ e^{-\Delta E} & \text{其它}\end{cases} \tag{4.6.10}$$

$\triangle E$ 为新组态 $\{x_j\}$ 与原组态 $\{x\}$ 的能量差。

以上过程循环多遍，系统温度降低得足够慢可以得到很好的恢复结果。以下实验的条件是信噪比 $SNR=1.25$，方差 $\sigma^2=40$，含噪图像中 $f$ 为图像的保真度，$f=\dfrac{M\times N-Ne}{M\times N}\times 100\%=\left(1-\dfrac{Ne}{M\times N}\right)\times 100\%$，$M\times N$ 为图像阵列点数，$Ne$ 为误差象元的总数，$K$ 为循环次数。

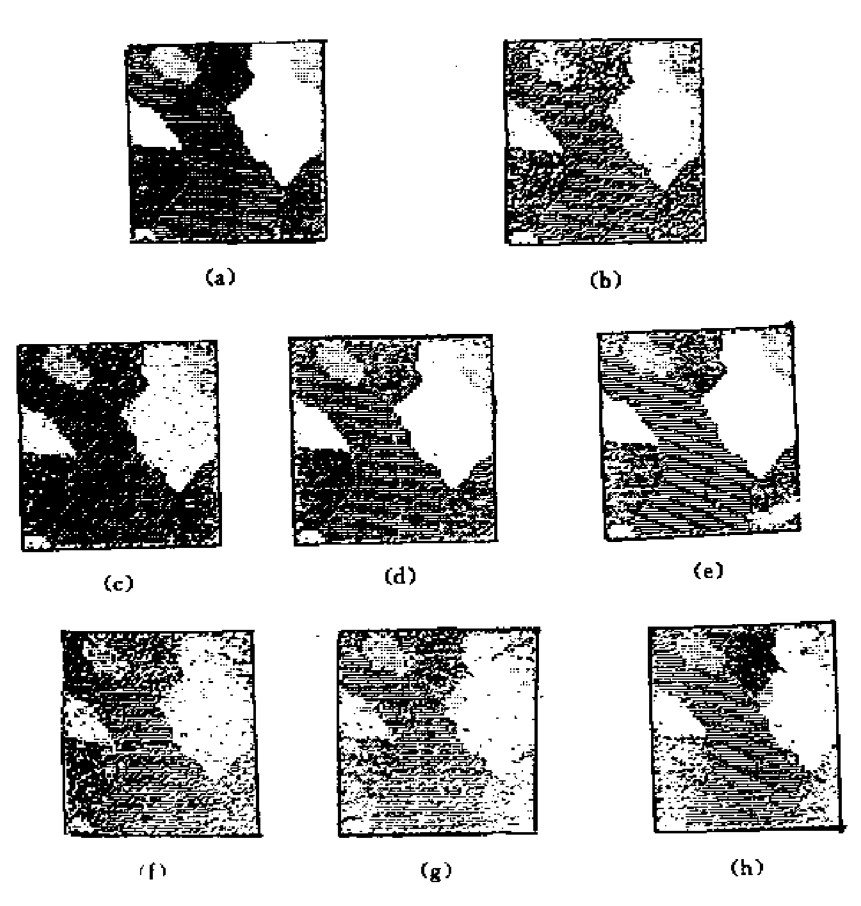


图 4-4　用模拟退火法进行图像恢复的结果

## 二、最大熵恢复

熵(entropy)是一个古老的概念，它表征平均信息量的大小，在数字图像处理和模式识别中也有很多的应用。如在图像编码中，用熵表示相应于编码器输入值集合的信息量，它提供了一个量度任何特定码性能的准则。如已知编码器的输入概率，则熵可作为这些输入值编码器所

需平均比特数的下限，编码器所需平均比特数 $R$ 接近于熵 $H$，则编码器最佳。在模式识别中往往需要估计表征随机变量的概率密度 $P(x)$，这可用在一定约束条件下的最大熵原理来估计 $P(x)$，所获得的平均信息量最大，偏差最小。

熵的一般定义为

$$H=-\int_{-\infty}^{\infty}P(x)\ln P(x)\mathrm{d}x \tag{4.6.11}$$

式中 $P(x)$ 为随机变量 $x$ 的概率密度。

对离散信号熵表示为

$$H=-\sum_{K=1}^{M}P(k)\ln P(k) \tag{4.6.12}$$

熵的概念是表征随机变量集合的随机程度的量度。最小随机情况是某一随机变量的概率为 1，那结果是预先知道的，其熵 $H=0$。最大熵的情况是所有随机变量等可能性，即 $P_1=P_1=\cdots=P_m=\frac{1}{M}$，其熵最大，最大熵为 $H=\ln M$。因此最大熵是处于 $0\sim\ln M$ 之间，此时 $P(x)$ 是 $0\sim1$ 之间的值，$H$ 不可能出现负值，故最大熵准则能自动地引向全正的输出结果。

在二维数字图像中，熵定义为

$$H_f=-\sum_{m=1}^{M}\sum_{n=1}^{N}f(m,n)\ln f(m,n) \tag{4.6.13}$$

此时熵的定义与式(4.6.12)略有不同，在式(4.6.12)中熵被定义为一个离散随机过程。而在二维数字图像中熵被定义为一个单一的确定的正值函数，可利用最大熵原理估计 $\hat{f}(x,y)$。

数字图像最大熵复原的基本原理是将 $f(x,y)$ 写成随机变量的统计模型，然后在一定的约束条件下，找出用随机变量形式表示的熵的表达式。再用求极大值的方法，获得最优估计解 $\hat{f}(x,y)$。最大熵复原的含义是对 $\hat{f}(x,y)$ 起最大平滑估计。据 70 年代以来所发表的资料，最大熵复原基本上有二种方法，弗里登法(Friend)和伯格法(Burg)。这二种方法基本原理相同，只是对模型的假设方法不同，导致了 $\hat{f}(x,y)$ 的不同解。

Friend 建议的 ME 方法，首先假定图像函数具有非负值，即

$$f(x,y)\geqslant 0 \tag{4.6.14}$$

对于一幅 $N\times N$ 大小的图像，其总能量 $E$ 是一个固定的数

$$E=\sum_{x=1}^{N}\sum_{y=1}^{N}f(x,y) \tag{4.6.15}$$

同时定义

$$H_f=-\sum_{x=1}^{M}\sum_{y=1}^{N}f(x,y)\ln f(x,y) \tag{4.6.16}$$

由于 $H_f$ 的结构类似于信息论中的熵的表示法，所以称它为图像的熵。类似地可定义噪声的熵 $H_n$，但考虑到图像中噪声值可正可负，为保证图像的非负性，定义噪声为

$$n'(m,n)=n(m,n)+B \tag{4.6.17}$$

这里 $B$ 为最大的噪声负值。于是 $H_n$ 可定义为

$$H_n=-\sum_{m=1}^{N}\sum_{n=1}^{N}n'(m,n)\ln n'(m,n) \tag{4.6.18}$$

ME 恢复就是在满足式(4.6.15)和图像退化模型的约束条件下使恢复后的图像熵和噪声熵达到最大。利用求条件极值的 Lagrange 乘子法，引入函数

$$R = H_f + \rho H_n + \sum_{m=1}^{N}\sum_{n=1}^{N}\lambda_{mn}\{\sum_{x=1}^{N}\sum_{y=1}^{N}h(m-x,n-y)f(x,y) + n'(m,n) -$$

$$\beta - g(m,n)\} + \beta\{\sum_{x=1}^{N}\sum_{y=1}^{N}f(x,y) - E\} \tag{4.6.19}$$

式中 $\lambda_{mn}(m,n=1,2\cdots N)$和 $\beta$ 是$(N^2+1)$个 Largrange 乘子，$\rho$ 是加权因子，用来强调 $H_f$ 和 $H_n$ 相互之间的份量，Friend 建议取 $\rho=20$。若 $\hat{f}(x,y)$和 $\hat{n}(m,n)$代表 $f(x,y)$和 $n'(m,n)$的估值，则必须有

$$\left.\frac{\partial R}{\partial f(x,y)}\right|_{f=\hat{f}}=0 \tag{4.6.20}$$

$$\left.\frac{\partial R}{\partial n'(m,n)}\right|_{n'=\hat{n}}=0 \tag{4.6.21}$$

把式(4.6.19)代入式(4.6.20)和式(4.6.21)，经运算和整理得到

$$\hat{f}(x,y)=e^{[-1+\beta+\sum_{m=1}^{N}\sum_{n=1}^{N}\lambda_{mn}h(m-x,n-y)]} \qquad (x,y=1,2,\cdots,N) \tag{4.6.22}$$

和

$$\hat{n}(m,n)=\exp(-1+\lambda_{mn}/\rho) \qquad (m,n=1,2,\cdots,N) \tag{4.6.23}$$

而且 $\hat{f}(x,y)$和 $\hat{n}(m,n)$满足下列约束条件

$$\sum_{x=1}^{N}\sum_{y=1}^{N}\hat{f}(x,y)=E \tag{4.6.24}$$

$$\sum_{x=1}^{N}\sum_{y=1}^{N}h(m-x,n-y)f(x,y)+\hat{n}(m,n)-\beta=g(m,n) \qquad (m,n=1,2\cdots,N) \tag{4.6.25}$$

式(4.6.22)即为恢复的图像函数。把式(4.6.22)和式(4.6.23)代入式(4.6.24)和式(4.6.25)可得$(N^2+1)$个方程，由此联立方 程组可解得$(N^2+1)$个未知数 $\lambda_{mn}(m,n=1,2\cdots N)$和 $\beta$，从而可逐点求得 $\hat{f}(x,y)$的值。只要 $\beta$ 取得足够大，保证 $n'(m,n)>0$，使用$(N^2+1)$维的 Nowton-Raphson 方法总能求得这$(N^2+1)$元联合方程组的解。

熵通常取决于 $f$ 的形状。当图像具均匀的灰度时熵最大，因此用 ME 恢复的图像就具有某种平滑性，前面提到一个函数的内积可用为该函数平滑性的测度，因此自然地也可以把内积函数用于 ME 恢复。

## §4.7 同态滤波复原

同态滤波复原与第三章介绍的同态图像增强技术相似，其基本原理是先对降质图像取对数，再进行滤波处理，最后通过指数变换得到复原图像 $\hat{f}(x,y)$。但它们使用的场合和目的不一样。

设退化图像 $g(x,y)$可以分为两部分乘积，即

$$g(x,y)=i(x,y)r(x,y) \tag{4.7.1}$$

取对数得

$$\lg g(x,y)=\lg i(x,y)+\lg r(x,y) \tag{4.7.2}$$

其复原过程如下




图 4－5　同态滤波器复原框图

设同态滤波器的冲激响应为 $I(x,y)$，其复原结果为 $\hat{f}(x,y)$，即

$$\hat{f}(x,y)=10^{[I(x,y)*(\lg I(x,y)+\lg r(x,y))]} \tag{4.7.3}$$

同态滤波在不考虑相位的情况下，也可用频域复原方法进行。其复原的准则是估计图像 $\hat{f}(x,y)$ 的功率谱 $S_{\hat{f}}(u,v)$ 与原图像 $f(x,y)$ 的功率谱 $S_f(u,v)$ 相等，即

$$S_{\hat{f}}(u,v)=S_f(u,v) \tag{4.7.4}$$

根据上述准则设计的同态复原滤波器 $L(u,v)$ 为

$$|L(u,v)|=\left[\frac{S_f(u,v)}{S_g(u,v)}\right]^{1/2} \tag{4.7.5}$$

式中 $S_g(u,v)$ 为降质图像的功率谱。由此前讨论的图像退化模型如下式所示

$$g=Hf+n \tag{4.7.6}$$

式中 $H$ 为降质系统的冲激响应，$n$ 为噪声。所对应的退化图像的功率谱为

$$\begin{aligned} S_g(u,v)&=E\{G(u,v)G(u,v)^*\} \\ &=E\{[H(u,v)F(u,v)+N(u,v)][H(u,v)F(u,v)+N(u,v)]^*\} \\ &=|H(u,v)|^2S_f(u,v)+S_n(u,v) \end{aligned} \tag{4.7.7}$$

式中 $H(u,v)$ 为降质系统的传递函数，$S_f(u,v)$ 和 $S_n(u,v)$ 分别为原始图像和噪声的功率谱。

将式(4.7.7)代入式(4.7.5)后可得同态滤波器的传递函数的表示式为

$$\begin{aligned} |L(u,v)|&=\left[\frac{S_f(u,v)}{|H(u,v)|^2S_f(u,v)+S_n(u,v)}\right]^{1/2} \\ &=\left[\frac{1}{|H(u,v)|^2+\dfrac{S_n(u,v)}{S_f(u,v)}}\right]^{1/2} \end{aligned} \tag{4.7.8}$$

同态滤波器的传递函数与维纳滤波器的形式，除分子相差一项 $H^*(u,v)$ 以外基本相似，其求解的关键还是要预先知道功率谱 $S_n(u,v)$ 和 $S_f(u,v)$。

如噪声项为零，其滤波器的传递函数为 $\dfrac{1}{H(u,v)}$，就是前面讨论过的逆滤波器。

## 习　题

4－1　如成象过程只有 $y$ 方向的匀速直线运动，其速率为 $y_0=b_t/T$，其中 $T$ 为曝光时间，$b$ 为像移距离。试求该运动引起的降质系统的传递函数 $H(u,v)$ 和相应的点扩展函数 $h(x,y)$。

4－2　设两个系统的点扩展函数都是 $h_1(x,y)$，其大小为

$$h_1(x,y)=\begin{cases} e^{-(x+y)} & x\geqslant 0 \quad y\geqslant 0 \\ 0 & \text{其它} \end{cases}$$

若将此两个系统串接，试求此系统的总的冲激响应 $h(x,y)$。

4－3 在连续线性位移不变系统的维纳滤波器中，如果假设噪声与信号的功率谱之比为 $S_n(u,v)/S_f(u,v)=|H(u,v)|^2$，试求最佳估值 $\hat{f}(x,y)$ 的表示式。

# 参 考 文 献


［4－1］ Andrews H.C. and Hunt B.R. Digital Image Restoration Prentice-Hall,1977

［4－2］ Phillips D.L.,A technique for the manerical solution of Certain integral equations of the first kind, J. Assoc Comp, March, Vol.9 1962.

［4－3］ 陈廷标等著.数字图像处理.北京:人民邮电出版社.

［4－4］ 张梅英,沈庭芝。用模拟退火法进行图像复原。北工学报,Vol.17,1997,P82

# 第五章　数字图像的压缩编码

## §5.1　概　　述

许多年以来为了研制电视和传真传输用的数字图像压缩编码系统已经经过了很大的努力，它的设计是要有一定要求的逼真度和使用尽可能少的代码比特数(bit)来表达图像，以压缩图像的存贮量，扩大传输容量，提高传输速度，从图像处理的研究中可以看到图像的象点之间有很大的相关性，因此它有很大的冗余性可以进行压缩。为此进行图像的编码与压缩是具有很大的潜力的。

我们这里指的图像信息都是数字化了的，它可以是物体的自然景象或称二维视觉图像，也可以是雷达的距离与速度数据阵列，它们可以是单色的、彩色的、多光谱的，也可以是二值的，如计算机图形，文字显示图像，也可以是连续色调的，如照相、印刷、电视图像等。但是，在数字化图像出现之前就已经有图像压缩编码方法的存在，其中最典型的是彩色电视系统。目前世界上有三种电视制式标准：美国的国家电视制式委员会制(NTSC)、西德的逐行倒相制(PAL)和法国的彩色顺序记忆制(SECAM)。

从彩色电视各种系统中可以知道，彩色电视信号都是利用其频谱的不连续性，把色度信号插到亮度信号频谱的低能量间隔区中，从而使彩色电视信号所占频谱几乎与黑白电视信号相同，并且实现了彩色和黑白电视的兼容性。这就是早期图像频带压缩工作的辉煌成果。

在这一章中主要讨论数字图像的压缩编码方法。数字化图像的传输有其显著的特点，那就是数字化图像的质量高，具有比较高的抗干扰性能，而数字化图像传输在未来的信息社会中，便于和数字电话、计算机数据传输合起来形成综合服务的信息网。由于数字计算技术的飞速发展，数字图像压缩技术在方法和实现上都有了更为广泛的领域。数字图像信息的压缩，在理论研究和具体实现上都已经有了许多尝试和实践，这里只能作一些重点的介绍，以了解重要情况。

数字图像压缩不但在传输上有其重要性，而且在图像数据的存贮方面也越来越显出其必要性。尤其是陆地卫星发射以后。陆地卫星是一颗地球资源卫星，它所获取的信息量非常大，每天要寻取 3000 多张图像，每张图为 2340×3240 个象素，每象素 6～8 bit，代表 185×185 $km^2$ 大小的一块地球表面。每张图有 4～7 个波段，因此每天要录取 11000 多张图，其有 $1\times10^{12}$ bit 左右的信息量。这些图像数据要保存几年供使用单位索取转录。如果能压缩一半，就可以少造一栋楼房，少用一半磁带，有很大的经济价值。当然信息量应尽可能少丢失。

图像信息压缩的可能性存在于图像本身之中，其中，明显可以利用的一点是图像各象素点之间的相关性。由于图像总是对一些物体的再现，而物体上各象素的灰度值之间有着极大的相关性；如把这些象素之间的关系信息提取出来，以一定的方法加以清除，那就可以压缩总的信息量，从统计的观点来说各象素的灰度分布远非独立的，图像信息的这种压缩潜力是很大的，当前所采用的各种压缩方法离压缩的极限还很远。这一点在以后还将讨论。

## §5.2 预测编码数字压缩技术

在卫星通信网和光纤通信网越来越广泛应用的当今时代，为了提高电视图像的质量，数字电视日益得到广泛的应用。在数字电视中，模拟电视信号和声音信号被采样和量化，并用脉码调制(PCM)方式传输。脉码调制有三步，即采样保持、量化和编码。采样频率按采样定理的要求必须大于信号频带宽度的两倍以防止混叠现象的发生。实际上，在电视系统中为补偿低通滤波器的响应，$F_S$ 至少高于通带截止频率 20%以上。例如对 NTSC 制，带宽为 $W=4.3$ MHz，因此取 $F_S=10.3$ MHz。另一方面，为了避免与彩色副载波频率 $F_{SC}$之间的可见拍频现象，$F_S$ 必须是 $F_{SC}$的谐波，NTSC 制 $F_{SC}=3.58$ MHz 其三次谐波为 10.74 MHz，四次谐波为 14.32 MHz。

对于一个 NTSC 制带宽为 4.3 MHz 的信号，经 $A/D$ 变换成为 8 bit 二进制数字信号，如采样频率取 10.74 MHz 则得出的二进制信息率为 86 Mbit/s。相对于模拟信号基带宽度 4.3 MHz 来说提高了 20 倍。由此可以看到数字图像传输中的压缩编码的必要性。数字图像压缩编码涉及量化、预测、变换、编码技术。这里首先从预测这一方面开始讨论这个问题。

预测编码系统把图像按行扫描进行编码。在扫到某一象素前，可以用此象素前面的一些象素值对其值进行预测估计，然后与实际象素值进行比较。即用实际值减去预测估值得到差值信号，再将此差值信号量化、编码和传输。在接收端则用量化的差值信号重建图像信号，其原理如图 5-1 所示。

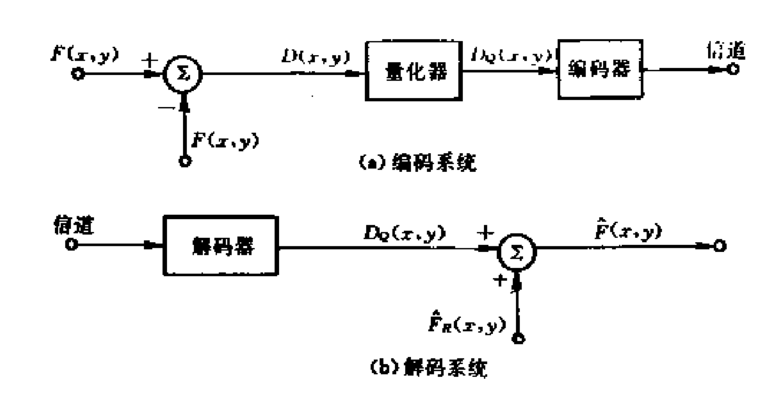



图 5-1 预测编码示意图

预测编码器的最简单的形式是 Δ 调制系统。如图 5-2 所示，将视频信号加到差分器的输入端，若差值信号是正的，脉冲发生器便在采样瞬间产生一个正脉冲，否则产生负脉冲。然后将这些正负脉冲表示为二进制码加以传输，同时将这些脉冲反馈到一个积分器上产生梯级形状的参考输入信号与实际输入信号相比较以产生差值信号。Δ 解调器则依次接收这些图像代码，并重建正负脉冲，然后对这些脉冲积分重建视频信号。

从图 5-2 中可以看到 Δ 调制系统的基本问题是：颗粒误差和斜率过载误差现象。由图中可以清楚地看到颗粒性误差是信号平坦区来回量化产生的，而如果为了减少量化误差而采取小量化跳步，那就不能精确地跟上快速上升的视频信号的斜率变化，除非采用很高的采样频率，否则两者不能兼顾。因此在固定的采样频率下，只能在颗粒误差和斜率过载之间取折衷。

对 Δ 调制的改进是差值脉冲编码调制(DPCM)。DPCM 图像压缩编码系统的方框图如图 5-3 所示。这一系统对实际象素值与其估值的差值进行量化和编码再进行传输。其估值是前

一象素的估值与差值信号之和，它应该就是前一象素值。对差值信号的量化分层，可采用不均匀分层，这样可把每象素 6～8 bit 的 PCM 压缩成 3 bit 的 DPCM。用 Huffman 变长度码时，可压缩到每象素 2.5 bit 的平均编码率，而不发生可以看得出的失真。

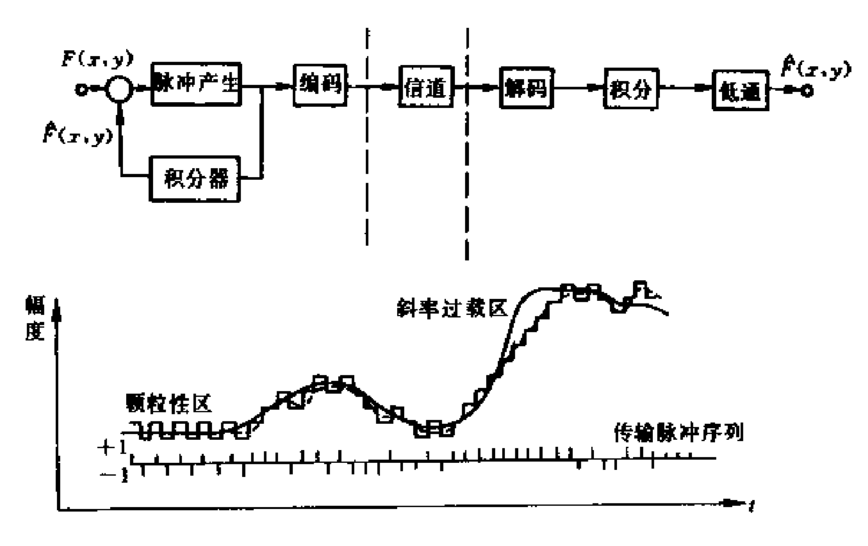



图 5-2 Δ调制编码系统

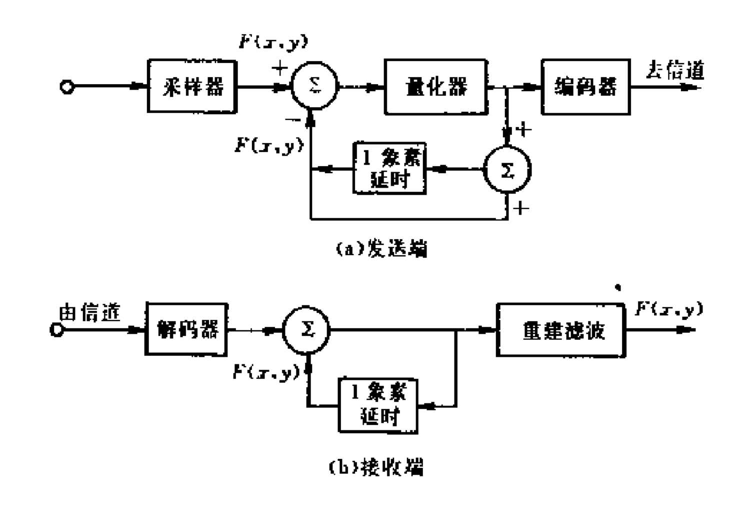



图 5-3 DPCM 调制系统

对图像数据的进一步压缩可以扩展 DPCM 的基本概念，就是用前几个扫描行邻近的象素来对当前象素进行估值，预测中利用更多的信息可以使均方误差减少并改善主观质量。令 $\hat{f}(m,n)$ 为预测值，令预测误差为

| | | | | | |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| | | $(m-1,n-1)$ | $(m,n-1)$ | $(m+1,n-1)$ | |
| 0 | 0 | 0 | 0 | 0 | 0 |
| | | $(m-1,n)$ | $(m,n)$ | | |
| 0 | 0 | 0 | 0 | | |

图 5-4 预测位置

$$e(m,n)=f(m,n)-\hat{f}(m,n) \tag{5.2.1}$$

$e(m,n)$又称为差值信号。由于 $\hat{f}(m,n)$与$f(m,n)$的邻近象素有关，即与$f(m,n)$存在相关性。因此只要以适当的方式、形式估值，其差值信号的相关性就大为减少。最佳估计 $\hat{f}(m,n)$往往是非线性估计，但为了计算上的方便往往用线性估计。例如用与$(m,n)$最近的三个象素来预测，可以写成

$$\hat{f}(m,n)=a_1 f(m-1,n)+a_2 f(m-1,n-1)+a_3 f(m,n-1) \tag{5.2.2}$$

选择系数 $a_1$、$a_2$ 和 $a_3$，使估计误差的均方值最小化，即

$$\min E\{[f(m,n)-\hat{f}(m,n)]^2\}$$

$$\begin{aligned}
&E\{[f(m,n)-\hat{f}(m,n)]^2\}\\
&=E\{[f(m,n)-a_1 f(m-1,n)-a_2 f(m-1,n-1)-a_3 f(m,n-1)]^2\}\\
&=E\{[f(m,n)^2+a_1^2 f^2(m-1,n)+a_2^2 f^2(m-1,n-1)+a_3^2 f^2(m,n-1)-\\
&\quad 2a_1 f(m,n)f(m-1,n)-2a_2 f(m,n)f(m-1,n-1)-\\
&\quad 2a_3 f(m,n)f(m,n-1)+2a_1a_2 f(m-1,n)f(m-1,n-1)+\\
&\quad 2a_1a_3 f(m-1,n)f(m,n-1)+2a_2a_3 f(m-1,n-1)\cdot f(m,n-1)]\}
\end{aligned}$$

令

$$\frac{\partial E}{\partial a_1}=0;\frac{\partial E}{\partial a_2}=0;\frac{\partial E}{\partial a_3}=0;$$

相关系数 $R(m,n;p,q)=E\{f(m,n)f(p,q)\}$得

$$\begin{aligned}
R(m,n;m-1,n)=&a_1R(m-1,n;m-1,n)+a_2R(m-1,n-1;m-1,n)+\\
&a_3R(m,n-1;m-1,n)\\
R(m,n;m-1,n-1)=&a_1R(m-1,n;m-1,n-1)+a_2R(m-1,n-1;m-1,n-1)+\\
&a_3R(m,n-1;m-1,n-1)\\
R(m,n;m,n-1)=&a_1R(m-1,n;m,n-1)+a_2R(m-1,n-1;m,n-1)+\\
&a_3R(m,n-1;m,n-1)
\end{aligned}$$

对于平稳随机场

$$R(m,n;p,q)=R(m-p,n-q)=R(\alpha,\beta)$$

对于 $f(m,n)$为平稳的一阶马尔可夫过程，有

$$R(\alpha,\beta)=R(0,0)\exp[-c_1|\alpha|-c_2|\beta|] \tag{5.2.3}$$

因此可得

$$\begin{aligned}
R(1,0)&=a_1R(0,0)+a_2R(0,1)+a_3R(1,1)\\
R(1,1)&=a_1R(0,1)+a_2R(0,0)+a_3R(1,0)\\
R(0,1)&=a_1R(1,1)+a_2R(1,0)+a_3R(0,0)
\end{aligned}$$

而

$$R(1,1)=\frac{R(1,0)R(0,1)}{R(0,0)}$$

因此解得

$$\alpha_1=\frac{R(1,0)}{R(0,0)};\quad \alpha_2=-\frac{R(1,1)}{R(0,0)};\quad \alpha_3=\frac{R(0,1)}{R(0,0)} \tag{5.2.4}$$

差值信号的均方值为

$$E\{e^2(m,n)\}=R(0,0)-[a_1R(1,0)+a_2R(1,1)+a_3R(0,1)] \tag{5.2.5}$$

由相关函数的性质可知 $R(1,1)$、$R(1,0)$和 $R(0,1)$都小于 $R(0,0)$，因此 $\alpha_1$、$\alpha_2$、$\alpha_3$ 都是绝对值小于 1 的值，把式(5.2.4)代入式(5.2.5)可知

$$E\{e^2(m,n)\}<R(0,0)$$

预测的值还可以进一步扩充到更大的邻域，如图 5-5 所示为一种预测邻域的方式。图中表示出各相应位置上的预测系数。其差值信号可表示为

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p_{3,3}$ | $p_{3,2}$ | $p_{3,1}$ | $p_{3,0}$ | $p_3q_3$ | $p_2q_3$ | $p_1q_3$ | $-q_3$ |
| $p_{2,3}$ | $p_{2,2}$ | $p_{2,1}$ | $p_{2,0}$ | $p_3q_2$ | $p_2q_2$ | $p_1q_2$ | $-q_2$ |
| $p_{1,3}$ | $p_{1,2}$ | $p_{1,1}$ | $p_{1,0}$ | $p_3q_1$ | $p_2q_1$ | $p_1q_1$ | $-q_1$ |
| $p_{0,3}$ | $p_{0,2}$ | $p_{0,1}$ | 1 | $-p_3$ | $-p_2$ | $-p_1$ | 1 |

图 5-5　预测系数的分布

$$e(m,n)=f(m,n)-\sum_{K=1}^{L}p_{k,0}f(m-k,n)-\sum_{K=1}^{L}p_{o,l}f(m,n-1)+\sum_{K=1}^{L}\sum_{l=1}^{L}p_{k,l}f(m-k,n-1) \tag{5.2.6}$$

式(5.2.6)中共有系数 $L^2+2L$ 个。当相关函数为可分离时，预测系数也是可分离的，这时差值信号可表示为

$$e(m,n)=f(m,n)-\sum_{k=1}^{L}p_Kf(m-k,n)-\sum_{l=1}^{L}q_lf(m,n-l)+\sum_{k=1}^{L}\sum_{l=1}^{L}p_kq_lf(m-k,n-l) \tag{5.2.7}$$

式(5.2.7)共有系数 $2L$ 个，式(5.2.7)可以用矩阵形式表示为

$$[e]=[Q][F][P] \tag{5.2.8}$$

式中的$[P]$和$[Q]$矩阵分别为

$$[P]=\begin{bmatrix} 1 & -p_1 & -p_2 & \cdots & -p_n & & 0 \\ & 1 & -p_1 & \cdots & \cdots & -p_n & \\ & & 1 & -p_1 & \cdots & \cdots & -p_n \\ & & & & & & \vdots \\ & & & & & & \vdots \\ & 0 & & & & & -p_1 \\ & & & & & & 1 \end{bmatrix} \tag{5.2.9}$$

$$
[Q]=\begin{bmatrix}
1 & & & & \\
-q_1 & 1 & & 0 & \\
-q_2 & & & & \\
\vdots & & & & \\
-q_n & \vdots & & & \\
 & -q_n & \cdots & 1 & \\
0 & & -q_n & \cdots & 1
\end{bmatrix}
\tag{5.2.10}
$$

因此预测压缩也可以看成是一种线性变换，但这种线性变换不是正交的。$e(m,n)$的分布由式(5.2.5)可知，要比原图像灰度分布更加集中，并且集中在0值附近，如图5-6所示。

现实的预测器可按图5-7所示的邻域次序编号得出，二阶的预测器用$S_1$和$S_3$或$S_1$和$S_2$值；三阶预测器用$S_1,S_2,S_4$或$S_1,S_3,S_7$；$S_1,S_2,S_3$值，四阶的可用$S_1,S_2,S_3,S_4$值，五阶器可用$S_1,S_2,S_3,S_4,S_5$值，六阶的可用$S_1,S_2,S_3,S_4,S_5,S_6$等值。预测压缩必须配以非均匀的量化器和编码来达到压缩的目的，量化器请参考其它有关书籍，而编码器将在后面讨论。

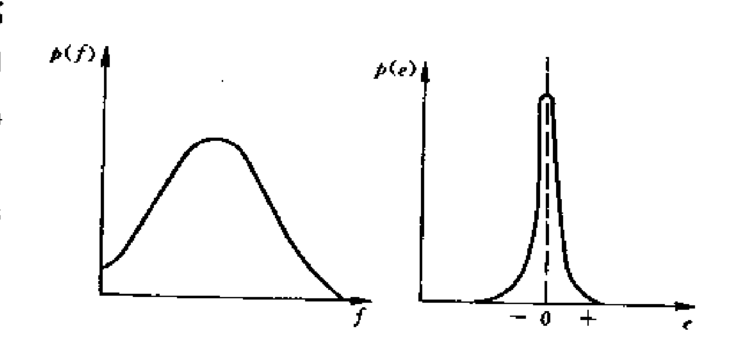
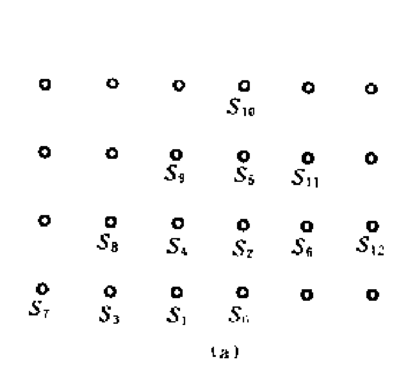



图5-6 分布密度函数



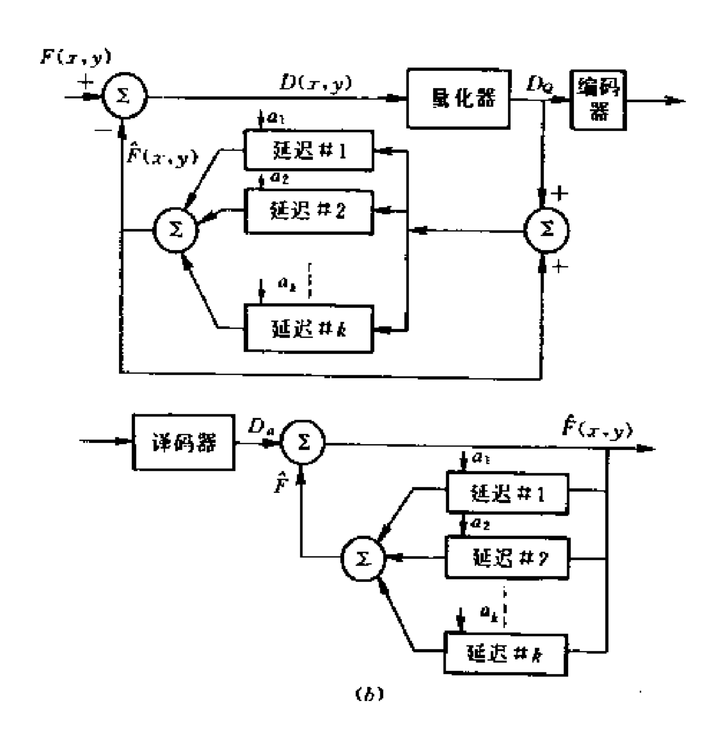




图5-7 高阶预测器示意图

预测压缩实现比较简单，但压缩能力不高而且抗干扰能力比较差，对传输中的误差有积累现象，以及在前面谈过的颗粒噪声和斜率过载现象。后者可以用自适应预测的办法来加以改进，即对物体边缘、平坦区和纹理区给予各不相同的预测函数，以减少预测误差。

## §5.3 变换编码数字压缩技术

如果不对图像数据本身进行编码，而对它的二维变换进行编码和传输也可以实现对图像的压缩传输，其基本概念是，大多数自然图像的变换函数有许多项的幅度很小，这些系数往往可以完全丢舍，这样就可以用很少的码位来编码，而引起的图像失真往往可以忽略不计。这些变换中能提供最小均方误差的是离散的 K－L 变换。

从消除图像象素之间的相关性来说，希望在一个图像经二维变换后各元之间变得互不相关，也就是每一个元都是不可预测的，这样可以预计得到最佳的能量集中。

一个二维图像函数 $f(x,y)$ 的正交函数展开为

$$\begin{aligned} f(m,n) &= \sum_{\mu=0}^{N-1}\sum_{v=0}^{N-1}F(\mu,v)\varphi^{(\mu,\nu)}(m,n) \\ &= \sum_{\mu=0}^{N-1}\sum_{\nu=0}^{N-1}F(\mu,v)P(m,\mu)Q(v,n) \end{aligned} \tag{5.3.1}$$

式中 $\varphi^{(\mu,\nu)}(m,n)$ 是正交函数集。当它为可分离时，可写成 $\varphi^{(\mu,\nu)}(m,n)=P(m,\mu)Q(v,n)$。这时式(5.3.1)可写成矩阵形式，$[f]$ 为 $N\times N$ 矩阵，$[F]$ 也一样

$$[f]=[P][F][Q] \tag{5.3.2}$$

或可重写成

$$[f] = \sum_{\mu=0}^{N-1}\sum_{v=0}^{N-1}F(\mu,v)[\varphi^{(\mu,\nu)}]$$

$f(m,n)$ 的自相关函数为

$$R(m,n;p,q)=E\{f(m,n)f(p,q)\}$$

如果 $E\{f(m,n)\}=0$ 而要求变换后的 $F(\mu,v)$ 各元函数不相关，则可得到

$$\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}R(m,n;p,q)\varphi^{(\mu,\nu)}(p,q) = \gamma_{\mu\nu}\varphi^{(\mu,\nu)}(m,n) \tag{5.3.3}$$

即此时的 $\varphi^{(\mu,\nu)}(p,q)$ 为自相关函数的特征函数，而

$$\gamma_{\mu,\nu}=E\{|F(\mu,\nu)|^2\}$$

式(5.3.2)的证明如下：由正交变换的特性得

$$F(\mu,\nu) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}f(m,v)\varphi^{(\mu,\nu)*}(m,n)$$

$$E\{F(\mu,\nu)\} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E\{f(m,n)\}\varphi^{(\mu,\nu)*}(m,n)$$

由于 $E\{f(m,n)\}=0$ 得 $E\{F(\mu,\nu)\}=0$。由于要求 $F(\mu,\nu)$是非相关的，因此

$$E\{F(\mu,\nu)F^*(\mu',\nu')\}=E\{F(\mu,\nu)\}E\{F^*(\mu',\nu')\}=0$$

由式(5.3.1)可得

$$\begin{aligned} E\{f(m,n)F^*(\mu',\nu')\} &= \sum\sum E\{F(\mu,\nu)F^*(\mu',\nu')\}\varphi^{(\mu,\nu)}(m,n) \\ &= E\{|F(\mu',\nu')|^2\}\varphi^{(\mu',\nu')}(m,n) \end{aligned}$$

另一方面由于

$$F^*(\mu',\nu') = \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} f(p,q)\varphi^{(\mu',\nu')}(p,q)$$

对上式两边乘 $f(m,n)$并取其期望值，得

$$E\{f(m,n)F^*(\mu',\nu')\} = \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} R(m,n;p,q)\varphi^{(\mu',\nu')}(p,q)$$

因此得

$$\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} R(m,n,p,q)\varphi^{(\mu',\nu')}(p,q) = E\{|F(\mu',\nu')|^2\}\cdot\varphi^{(\mu',\nu')}(m,n)$$

证毕。

上述式(5.3.2)也可表示成矩阵形式，把[$f$]写成列串接矢量形式 $f$，$f$ 为 $N^2\times 1$ 矢量，则基函数矩阵[$\varphi^{(\mu,\nu)}$]也可写成矢量形式 $\Phi^s$，$s$ 即为$(\mu,\nu)$之和。这时自相关函数 $R(m,n;p,q)$可写成[$k(i,j)$]为 $N^2\times N^2$ 矩阵，$i$ 代表$(m,n)$，$j$ 代表$(p,q)$。$k(i,j)=E\{f_i f_j\}$，$(i=0,1,\cdots,N^2-1;j=0,1,\cdots,N^2-1)$。特征方程为

$$\left.\begin{aligned} &\sum_{j=0}^{N^2-1} k(i,j)\varphi_j^s = \gamma_s\varphi_i^s \\ &i=1\cdots N^2-1 \end{aligned}\right\} \text{共有 } N^2 \text{ 个} \qquad (S=0,\cdots,N^2-1)$$

$\varphi^S=[\varphi_0^S\varphi_1^S\cdots\varphi_{N^2-1}^S]^T$ 为[$k$]的特征矢量，$\gamma_S$ 为特征值，因此 $\varphi^S$ 使[$f$]变换成由互不相关元组成的[$F$]。当自相关函数为可分离时

$$[k]=[k_y]\otimes[k_x]$$

$$[k_x]=[\Phi_x][\Lambda_x][\Phi_x]^t; \qquad [k_y]=[\Phi_y][\Lambda_Y][\Phi_y]^t$$

[$\Lambda_x$]和[$\Lambda_y$]为 $N\times N$ 对角矩阵，则

$$[F]=[\Phi_x]^t[f][\Phi_y] \tag{5.3.4}$$

K－L 变换可以保证变换后各元互不相关，在丢舍一些小值项后能保证均方误差最小，但它要求知道图像的统计参数，即要求知道被变换图像的自相关矩阵，而且没有快速算法。经近年来研究证明，余弦变换的性能与 K－L 变换相近似，并有快速算法。当图像的自相关函数是可分离的且是一阶马尔可夫过程时，相关函数可以表示成

$$E[f(m,n)f(p,q)]=e^{-\alpha|m-p|}e^{-\beta|n-q|}$$

这时的 K－L 变换具有快速算法，$\varphi^{(\mu,\nu)}(m,n)$为正弦函数

$$\varphi^{(\mu)}(m)=\sqrt{\frac{2}{N+\gamma_\mu}}\sin\left[\omega_n\left(m-\frac{N-1}{2}\right)+\frac{\mu+1}{2}\pi\right] \tag{5.3.5}$$

式中

$$\gamma_\mu=\frac{1-\alpha^2}{1-2\alpha\cos\omega_n+\alpha^2}$$

在第二章中所述的其它各种二维变换，如傅氏变换、余弦变换、哈达玛变换、沃尔什变换等都有其计算上的特点，而且都是采用固定的变换基，不必知道图像的统计参数。把各种变换的公式和计算量列出表 5－1。

**表 5－1　各种变换列表**

| 变换名称 | 表示式 | 未知参数 | 计算量 | 注 |
|---|---|---|---|---|
| 1. K－L 变换 | $[K_x]=[\Phi_x][\Lambda_x][\Phi_x]$<br>$[K_y]=[\Phi_y][\Lambda_y][\Phi_y]^t$<br>$[F]=[\Phi_x]^t[f][\Phi_y]$ | 变换系数 | $N^3\rightarrow 2N^3$ | 当图的统计参数固定时，只须计算一次 |
| 2. 傅氏变换 | $[F]=[W][f][W]$ | 变换系数 | $2N^2\log_2N$<br>（复数） | 当 $N\neq 2K$ 时，计算量增加 |
| 3. 余弦变换 | $[F]=[C]^t[f][C]$ | 变换系数 | $2N^2\log_2N$<br>（实数） | 同上 |
| 4. 哈达玛变换 | $[F]=[H][f][W]$ | 变换系数 | $2N^2\log_2N$ | 当 $N\neq 2^K$ 时，计算量增加 |
| 5. 预测变换 | $[e]=[P][f][Q]$ | | | |

变换后的图像的编码方法有两种，即门限编码法和区域编码法。

1. 区域编码

图像矩阵$[f]$经二维变换后的 $N\times N$ 二维矩阵$[F]$表现出能量集中，但是 $N$ 很大时，进行二维变换的计算量是很大的，实际进行的是用分块变换的办法，即把$[f]$分为 $n\times n$ 的许多小块。每一块不能太小，也不必太大，一般用 16×16 方块即可。分块后的问题是吉布斯效应，它主要发生在每一块的边缘象素处。区域编码就是对一个区域中变换后的 16×16 个系数中的每一系数给以不同的二进码位数。例如典型的码位赋值如图 5－8 所示。

| 8 | 8 | 8 | 7 | 7 | 7 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 8 | 7 | 6 | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 |
| 8 | 7 | 6 | 4 | 4 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 7 | 6 | 4 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 5 | 4 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 5 | 4 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 3 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 3 | 3 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 3 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

图 5-8　区域变换编码典型码位赋值图

对变换后的一半以上的元赋以 0 位，即舍弃之。当然，对各种变换可以赋以不同的码位分布，所得到的均方误差也不同。按图 5-8 表示的码位分布为平均每象素 1.5 码位。

2. 阈值编码(门限编码)

在阈值编码方法中，变换后各元的值如超过某一给定的阈值时给以一定的量化层数，并对其幅值编码。对变换的每个元的位置都必须在编码后传输到接收端。一种比较简单的方法是对阈值以下丢弃的无效样本数目进行编码，这也是一种行程长度编码方案。由于这种阈值编码是自适应的，即根据各个不同图像选取不同的传输元，因此，它的性能多少要比区域编码好些。但在阈值固定的情况下，传输系数的数目和码位数是与图像有关的。这对固定传输率的通信干线是不能适应的。因此，有过一种固定码率而自适应地选取阈值的方法。

在图像变换编码中，主要的压缩技术是只传输变换域中某些系数，而把一些幅值小的系数丢弃掉，在接收端对这些丢弃的系数以零代替。也可以在接收端采用频谱外插的方法对这些系数进行估计。实践证明用外插估计的方法对图像质量的改善是明显的。变换压缩的优点是压缩比比较大，抗干扰能力强。因为干扰成分在接收端反变换后分散到全图中去了，对每一个象素的影响不会太大。传输中产生的误差也不会积累。变换压缩的缺点是计算复杂，还难以实时实现。采用变换编码和 DPCM 预测编码相结合的办法可以得到比较好的结果。例如对 $N\times N$ 图像阵列 $f(m,n)$ 的每一行作一维变换，然后对变换系数进行预测编码，即令

$$F_m(v) = \sum_{n=0}^{N-1} f(m,n)A(n,v)$$

DPCM 产生的差值信号为

$$d(v)=F_m(v)-\hat{F}_m(v)$$

而系数估值是由前一行的系数加权形成的

$$\hat{F}_m(v)=a_m\hat{F}_{m-1}(v)$$

变换核可以采用多种形式，如余弦变换、傅氏变换、Hadamard 变换等等。

## §5.4 熵编码方法

### 一、一些基本概念

1. 图像熵和平均码字长度

(1)图像熵 设图像象素灰度级集合为$(w_1,w_2,\cdots,w_k,\cdots,w_M)$，其对应的概率分别为$P_1,P_2,\cdots,P_K,\cdots,P_M$。按信息论中信源信息熵定义，数字图像的熵$H$为

$$H=-\sum_{k=1}^{M}P_k\log_2 P_k \qquad \text{(bit)} \tag{5.4.1}$$

由此可见，图像熵$H$是表示各个灰度级比特数的统计平均值。

(2)平均码字长度 设$B_K$为数字图像第$K$个码字$C_K$的长度(二进制代数的位数)。其相应出现的概率为$P_K$，则该数字图像所赋于的码字平均长度$R$为

$$R=\sum_{k=1}^{m}B_K P_K \qquad \text{(bit)} \tag{5.4.2}$$

(3)编码效率 在一般情况下，编码效率往往用下列简单公式表示

$$y=\frac{H}{R}(\%) \tag{5.4.3}$$

式中 $H$为信源熵，$R$为平均码字长度。

根据信息论中信源编码理论，可以证明在$R\geqslant H$的条件下，总可以设计出某种无失真编码方法。当然如编码结果使$R$远大于$H$，表明这种编码方法效率很低，占用比特数太多。例如对图像样本量化值直接采用PCM编码，其结果平均码字长度$R$就远比图像熵$H$大。最好编码结果使$R$等于或很接近于$H$。这种状态的编码方法，称其为最佳编码，它既不丢失信息而不引起图像失真，又占用最少的比特数。例如下面要介绍的哈夫曼编码即属最佳编码方法。若要求编码结果$R<H$，则必然丢失信息而引起图像失真，这就是在允许失真条件下的一些失真编码方法。

熵编码的目的就是要使编码后的图像平均比特数$R$尽可能接近图像熵。一般是根据图像灰度级出现的概率大小赋于不同长度码字，概率大的灰度级用短码字，反之，用长码字。可以证明，这样的编码结果所获得的平均码字长度最短。这就是下面要介绍的变长最佳编码定理。

2. 变长最佳编码定理

[定理] 在变长编码中，对出现概率大的信息符号赋于短码字，而对于出现概率小的信号符号赋于长码字。如果码字长度严格按照所对应符号出现概率大小逆序排列，则编码结果平均码字长度一定小于任何其它排列方式。

以上定理就是下面的哈夫曼编码方法的理论基础。通过对其证明的介绍进一步加强认识。

设图像灰度级为$W_1,W_2,\cdots,W_i,\cdots,W_N$；各灰度级出现概率分别为$P_1,P_2,\cdots,P_i,\cdots,P_N$；编码所赋于的码字长度分别为$t_1,t_2,\cdots,t_i,\cdots,t_N$，则编码后图像平均码字长度$R$应为

$$R=\sum_{i=1}^{N}P_i t_i \tag{5.4.4}$$

再令严格按照定理规则进行编码，其结果平均码字长度为$R_1$，$R_2$为将其中任两个灰度级不按

定理规则编码(即概率大的灰度级赋于长码字。反之,用短码字),而其它所有灰度级按定理规则编码所得图像的平均码字长度。那么 $R_2$ 应等于 $R_1$ 加上"不按定理规则编码所增加的平均码字长度"$\Delta R$。只要证明 $\Delta R$ 是大于零,即可以证明上述定理。

令第 $m$ 和 $n$ 个灰度级出现的概率分别为 $P_m$ 和 $P_n$,且 $P_m>P_n$。与这两个灰度级对应的码字长度分别为 $t_m$ 和 $t_n$。如果不按定理规则赋于这两个码字长度,即 $t_m>t_n$,那么演算下式

$$\begin{aligned}&(P_mt_m+P_nt_n)^2-(P_mt_n+P_nt_m)^2\\&=P_m^2t_m^2+P_n^2t_n^2+2P_mP_nt_mt_n-P_m^2t_n^2-P_n^2t_m^2-2P_mP_nt_mt_n\\&=P_m^2t_m^2+P_n^2t_n^2-P_m^2t_n^2-P_n^2t_m^2\\&=(P_m^2-P_n^2)(t_m^2-t_n^2)\end{aligned}$$

根据前面假设条件 $P_m>P_n$;$t_m>t_n$ 可知这个演算式结果一定大于零。将开始式展开立即可以得到

$$(P_mt_m+P_nt_n)-(P_mt_n+P_nt_m)=\Delta R>0$$

由于上式左边第一项是对这两个灰度级不按定理规则编码获得的码字长度统计平均值,而第二项是按照定理规则编码获得的码字长度统计平均值。两项之差当然就是前面定义的 $\Delta R$ 了。到此就可以直观地证明了变长编码定理。

3. 可变长度最佳编码的平均码字长度

设可变长度编码所用码元进制为 $D$,被编码的信息符号总数为 $N$,第 $i$ 个符号出现的概率为 $P_i$,与其相对应的码字长度为 $t_i$,则可以证明,这种编码结果平均码字长度 $R$ 落在下列区间内。

$$\frac{H}{\lg D}\leqslant R\leqslant\frac{H}{\lg D}+1$$

式中

$$H=-\sum_{i=1}^{N}P_i\lg P_i$$

由此可以引导出对某一信息符号存在下式

$$-\frac{\lg P_i}{\lg D}\leqslant t_i<-\frac{\lg P_i}{\lg D}+1 \tag{5.4.5}$$

对二进制码可以进一步简化为

$$-\log_2P_i\leqslant t_i<-\log_2P_i+1 \tag{5.4.6}$$

可见,码字的长度是由信息符号出现的概率来决定的,这就是下面要介绍的香农编码方法的理论基础。

4. 唯一可译编码

有些情况下,为了减少表示图像的平均码字长度,往往对码字之间不加同步码,但是,这样就要求所编码字序列能被唯一地译出来。满足这个条件的编码称其为唯一可译编码。也常称为单义可译码。单义可译码往往是采用非续长代码。

(1)续长代码和非续长代码　若代码中任何一个码字都不是另一个码字的续长,也就是不能在某一码字后面添加一些码元而构成另一个码字,称其为非续长代码。反之,称其为续长代码。如二进制代码[0,10,11]即为非续长代码,而[0,01,11]则为续长代码。因为码字 01 可由

码字“0”后加一个码元“1”构成。

(2)单义代码　任意有限长的码字序列，只能被唯一地分割成一个个码字，则这样的码字序列称为单义代码。单义代码的主要条件是满足克劳夫特(Kraft)不等式。

$$\sum_{i=1}^{n} D^{-t_i} \leqslant 1 \tag{5.4.7}$$

式中 $D$ 为代码中码元种类，对二进制 $D=2$；$n$ 为代码中码字个数；$t_i$ 为代码中第 $i$ 个码字的长度(即码元个数)。

如代码 $C=[00,10,001,101]$，因为是二进制码，则 $D=2$ 共有 4 个码字 $c_1=00,c_2=10,c_3=001,c_4=101$。其相应的长度为 $t_1=2,t_2=2,t_3=3,t_4=3$。代入式(5.4.7)可得

$$\sum_{i=1}^{4} 2^{-t_i} = \frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^3} = \frac{6}{8} < 1$$

因此可以证明代码 $C$ 是单义代码。

## 二、哈夫曼(Huffman)编码方法

哈夫曼编码是根据可变长度最佳编码定理，应用哈夫曼算法而定的一种编码方法。它的平均码字长度在具有相同输入概率集合的前提下，比其它任何一种唯一可译码都小。因此，也常称其为紧凑码。下面以一个具体例子说明其编码方法。

### 1. 编码步骤

(1)将输入灰度级按出现的概率由大到小顺序排列(对概率相同的灰度级可以任意颠倒排列位置)。

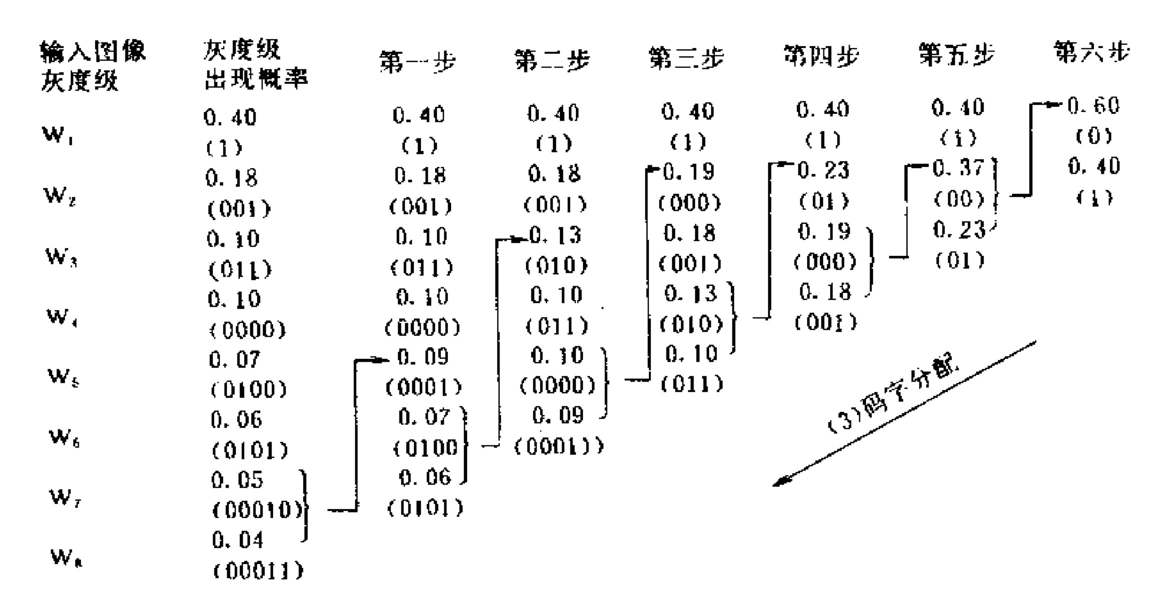
(2)将最小两个概率相加，形成一个新的概率集合。再按第(1)步方法重排(此时概率集合中，概率个数已减少一个)。如此重复进行直到只有两个概率为止。

(3)分配码字。码字分配从最后一步开始反向进行，对最后两个概率一个赋于“0”码，一个赋于“1”码。如概率 0.6 赋于“0”码，0.4 赋于“1”码。(也可以将 0.6 赋于“1”码。0.4 赋于“0”码)。如此反向进行到开始的概率排列。在此过程中，若概率不变仍用原码字。如表中第 6 步中概率 0.4 列第五步中仍用“1”码。若概率分裂为两个，其码字前几位码元仍用原来的。码字的最后一位码元赋于“0”码元，另一个赋于“1”码元。如表中第六步中概率 0.6 到第五步中分裂为 0.37 和 0.23，则所得码字分别为“00”和“01”。

### 2. 前例哈夫曼编码的编码效率计算

根据式(5.4.1)求出前例信源熵为

$$\begin{aligned} H &= -\sum_{i=1}^{8} P_i \log_2 P_i \\ &= -(0.41\log_2^{0.4} + 0.18\log_2^{0.18} + 2\times 0.1\log_2^{0.1} + 0.07\log_2^{0.07} + 0.06\log_2^{0.06} + \\ &\quad 0.05\log_2^{0.05} + 0.04\log_2^{0.04}) \\ &= 2.55 \end{aligned}$$




根据式(5.4.2)求出平均码字长度为

$$R=\sum_{i=1}^{8}B_iP_i=0.40\times1+0.18\times3+0.1\times3+0.10\times4+$$
$$0.07\times4+0.06\times4+0.05\times5+0.04\times5$$
$$=2.61$$

根据(5.4.3)式求出编码效率 $\eta$ 为

$$\eta=\frac{H}{R}=\frac{2.55}{2.61}=97.8\%$$

### 三、香农(Shannon)编码法

香农编码方法根据式(5.4.4)、式(5.4.5)，按下列步骤进行：

(1)将输入灰度级(信息符号)按出现的概率由大到小顺序排列(相等者可以任意颠倒排列位置)。

(2)按式(5.4.5)或式(5.4.6)计算各概率对应的码字长度 $t_i$。

(3)计算各概率对应的累加概率 $a_i$，即

$$a_1=0$$
$$a_2=P_1$$
$$a_3=P_2+a_2=P_2+P_1$$
$$a_4=P_3+a_3=P_3+p_2+P_1$$
$$\vdots$$
$$a_i=P_{i-1}+P_{i-2}+\cdots+P_i$$

(4)把各个累加概率由十进制转换成二进制数。

(5)将二进制表示的累加概率去掉多于(2)步中计算的 $t_i$ 的尾数，即获得各个信息符号的

码字。

为了能和哈夫曼码比较，我们仍用前例进行香农编码。具体步骤和结果如下：

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| 输入图像灰度级 | 灰度级出现概率 | 计算 $t_i$ | 计算 $a_i$ | 由十进制变为二进制 | 去掉多于 $t_i$ 尾数(码字) |
| $W_1$ | 0.40 | 2 | 0 | 0 | 00 |
| $W_2$ | 0.18 | 3 | 0.40 | 01100 | 011 |
| $W_3$ | 0.10 | 4 | 0.58 | 10010 | 1001 |
| $W_4$ | 0.10 | 4 | 0.68 | 10100 | 1010 |
| $W_5$ | 0.07 | 4 | 0.78 | 11000 | 1100 |
| $W_6$ | 0.06 | 5 | 0.85 | 1101100 | 11011 |
| $W_7$ | 0.05 | 5 | 0.91 | 1110100 | 11101 |
| $W_8$ | 0.04 | 5 | 0.96 | 1111010 | 11110 |

平均码字长度 $R$ 为

$$R = \sum_{i=1}^{8} p_i t_i = 0.4 \times 2 + 0.18 \times 3 + 0.10 \times 4 + 0.10 \times 4 + 0.07 \times 4 + 0.06 \times 5 + 0.05 \times 5 + 0.04 \times 5 = 3.17$$

计算其编码效率 $\eta$

$$\eta = H/R = 2.55/3.17 = 80.4\%$$

可见香农编码效率比哈夫曼编码效率低些，但还算是一种高效编码方法。

## §5.5 轮廓编码

### 一、轮廓编码的概念

一幅图像中总是存在许多大小不等的灰度级相同的区域，尤其是一些所谓特写图像，或某些几何图案的物体的照片等由少数恒定灰度级区域组成的图像。这类图像经过数字化以后，所得到的数字图像同样存在着少数灰度级量化级相同，大小不等的区域。假若我们能够将这些灰度级相同的区域从图像中找出来并给予不同的标志，那么我们只要对能够唯一确定这些区域的一些因素进行编码，也就等于对整个图像进行编码了。很明显，唯一确定某一区域只要三个因素：

(1)包围这个区域的外围边界即轮廓的方向序列。

(2)轮廓的起始点位置(行和列数)。

(3)轮廓所包围区的灰度级值。

可以看出，对每一个轮廓的三个因素编码，要比对轮廓包围区内每个象素都分配以码字节约很多比特数，而且图像细节越少，采用轮廓编码节省的比特数越多。

有三幅细节不同的图像。一幅是女孩面部特写，图像细节少。另一幅是摄影师半身工作环境照片，其图像细节中等。还有一幅是群众大会场面，照片有很多人头像，具有丰富的图像细

节。通过这三幅不同细节图像采用轮廓编码结果见表 5-2。

表 5-2

| 原灰度级数 | 比特数/象素 | 轮廓编码后图像的平均比特数 | | |
|---|---|---|---|---|
| | | 女孩 | 摄影师 | 群众场面 |
| 16 | 4 | 1.4 | 2.1 | 3.8 |
| 32 | 5 | 2.4 | 3.5 | 4.9 |
| 64 | 6 | 3.5 | 4.9 | 5.8 |
| 128 | 7 | 4.6 | 5.8 | 6.7 |

## 二、轮廓的算法

对图像要进行轮廓编码，首先得找出图像中的轮廓，用来寻找轮廓的方法 称为轮廓算法。

轮廓算法由两部分组成，一个是计算轮廓方向序列的方法称为 $T$ 算法；另一个是计算轮廓起始点的方法称为 IP(Initial Point)算法。

轮廓算法的步骤是由这两个算法依次交叉进行。即第一个轮廓起始点找到后，进行第一个轮廓方向序列的计算，算完后再寻找第二个轮廓起始点，接着计算第二个轮廓方向序列；如此依次交叉地进行，计算完图中所有轮廓为止。

下面分别介绍一下轮廓方向序列及轮廓起始点算法。

1. 轮廓方向序列的计算——T 算法

轮廓方向序列就是由轮廓起始点开始到轮廓上的第二点，第三点，…，直到最后一点再返回起始点为止方向所组成的方向序列。

确定轮廓上一点走向下一点的方向是用“最先左看规则”，即从进入轮廓点(如 $A$ 点)的方向看去(如图 5-9)，最先向左方向寻找，若遇到灰度级和 $A$ 点相同的邻点，则轮廓由 $A$ 点走向这一点，若左看没有灰度级和 $A$ 点相同的邻点，再按向上看、向右看、向下看的顺序寻找，直到找到有灰度级相同的点时，将轮廓由 $A$ 点移向该点。若四个方向都没有，表示这个轮廓只是由一个象素构成的。

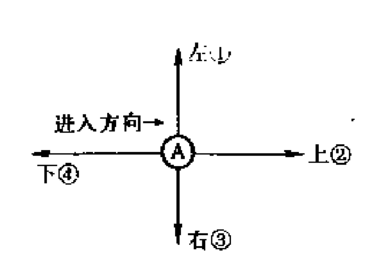



图 5-9 最先左看规则

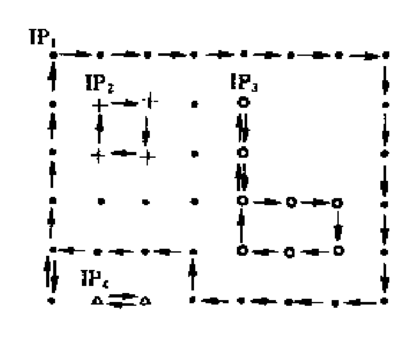



图 5-10 T 算法例图

例 5-1 应用“先左看”准则，(假定四个轮廓起始点 $IP_1, IP_2, IP_3, IP_4$ 都为已知)计算图 5-11 中四个轮廓走向序列 $T_1, T_2, T_3, T_4$。

2. 轮廓方向序列的标记

(1)一个轮廓方向序列应用“最先左看规则”找出来以后，这个轮廓就被确定了。对轮廓上

的每一个点，根据进入和离开的方向再按所谓“方向标记规则表”给予不同的标记。方向标记规则表见表 5-3。

表 5-3　方向标记规则表

| | | 离开轮廓点方向 | |
| --- | --- | --- | --- |
| | | ↑→ | ↓← |
| 进入轮廓点方向 | ↑← | *A* | *R* |
| | ↓→ | *R* | *D* |

例如，用“方向标记规则表”标记图 5-11 中的轮廓方向序列，见图 5-12。

(2)若遇一个轮廓点有两对进出方向时，即按“方向标记规则表”标记后有两个标志，再按“合并规则表”合并两个标志为一个标志，见表 5-4。

表 5-4　合并规则表

| 第一次、第二次通过的方向标记 | *DA* *AD* *RR* | *DR* *RD* *DD* | *AR* *RA* *AA* |
| --- | --- | --- | --- |
| 合并后的标记 | *R* | *D* | *A* |

例 5-2　按“合并规则表”，将图 5-11 中标志合并，见图 5-12。

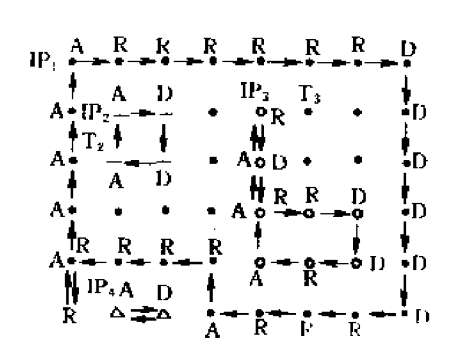


图 5-11　“方向标记”例图

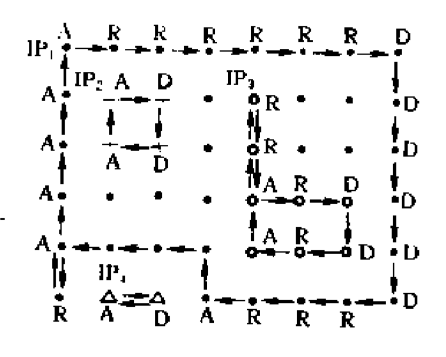


图 5-12　合并规则例图

(3)不是轮廓上的象素一律给予 *I* 标志，可见当数字图像标记完毕后存进贮存器时，要对每一象素额外分配两个比特来表示其标志符号 *A*、*D*、*R*、*I*。

## 三、轮廓起始点的寻找——IP 算法

为了寻找轮廓起始点 IP，我们使用顺序扫描搜索方法，也就是从数字图像的左角上(第一行，第一列)的象素开始，按行从左到右，按列从上到下逐点顺序扫描，直到右下角最后一个象素为止，对扫描遇到的每一个象素，进行判别是否为轮廓起点 *IP*。如何判别每一象素是否为 *IP* 呢？先来介绍一下“扫描搜索比较表”和“起始点判别准则”。

1. 扫描搜索比较表构成规则(CPL)

(1)每扫一行制一个表，当然开始扫描前表是空的。

(2)对每一行扫描从左到右逐点进行判别。若遇到点标志为 *A*，将该点的灰度级值从表的

右端依次填入表中。

若遇到点标志是 $D$,将表中最靠边标志为 $A$ 的灰度级划去。若遇到点标志是 $R$ 或 $I$,表不变化,即不填也不划去什么。

(3)每一行扫描完毕,表也一定是空的,因为轮廓总是封闭的,轮廓通过某一行有向下的点($D$),必有向上的点($A$),即 $A$ 点数一定等于 $D$ 点数。

2. 轮廓起始点判别准则

在扫描搜索过程中,凡是符合下列两个条件的就判定为轮廓起始点 IP。

(1)它的标志是 $I$(即不是已确认过的轮廓上的点)。

(2)它的灰度级值不等于扫描搜索表中最靠近的标记为 $A$ 的点的灰度级值。

这里要注意的就是一幅数字图像的左上角的点(第一行第一列)总认为是第一轮廓起始点,这是不难理解的。

## 四、应用轮廓算法的一个例子

若给定一幅数字图像如图 5 - 13 所示,对其进行轮廓算法的步骤如下:

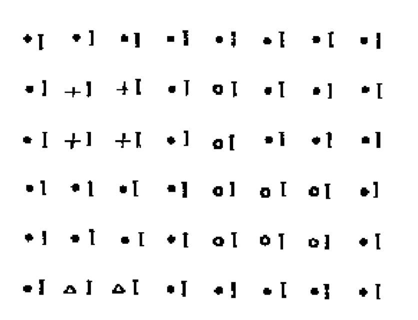


图 5 - 13　轮廓编码例图

图 5 - 14　轮廓编码例图

(1)先将图像各象素都标志为 $I$,如图 5 - 13。

(2)图像左上角点(第一行,第一列)总认为 $IP$。有了 $IP_1$ 根据"最先左看规则"找出第一轮廓走向序列 $T_1$,同时按"方向规则表"以及"合并规则表"将轮廓 $T_1$ 上各点标志改为 $A$ 或 $D$ 或 $R$ 等,见图 5 - 14。

(3)按 IP 算法寻找第二个轮廓起始点 $IP_2$。

第一扫描行的搜索表为:扫描的第一点标志为 $A$,其灰度级为"0"填入表的末尾。扫描第 2、3、4、5、6 标志为 $R$,不填入表内,即表不变。第 8 点标志为 $D$,将表中刚才填入的"0"划去。可见第一行扫完后,CPL 表是空的,表内没有起始点。

第二行扫描搜索表为:扫描的第一标志为 $A$,其灰度级为"0"填入表的末尾。扫描第二标记为 $I$,它的灰度级为"+",与表中相邻的标记为 $A$ 的灰度级值"0"不同,因此确定为新的轮廓起始,记为 $IP_2$。

再按 $T$ 算法找出 $T_2$,并标记好,见图 5 - 15。

再按 IP 算法找出 $IP_3$,进而找出 $T_3$,如此继续下去,直到找完全图为止,见图 5 - 16。

## 五、编码方法

一幅数字图像轮廓全部找出并标记完毕后,尚需给各轮廓分配码字。假设:

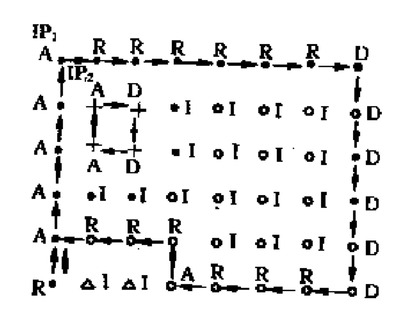



图 5-15　轮廓编码例图

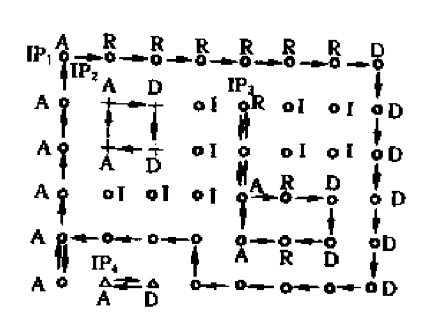



图 5-16　轮廓编码例图

(1)轮廓起始点位置(行和列数)用自然码;

(2)轮廓包围区的灰度级值也用自然码(也可用哈夫曼码);

(3)轮廓方向序列用链码表示。

根据上述假设,我们作下列约定,见表 5-5 对图 5-16 进行分配码字。

表 5-5

| 轮廓号 | 码字<br>自然码 | 灰度级 | 码字<br>自然码 | 行或列 | 码字<br>自然码 | 方向 | 码字<br>链码 |
|---|---|---|---|---|---|---|---|
| 1 | 00 | · | 00 | 1 | 000 | ↑ | 00 |
| 2 | 01 | + | 01 | 2 | 001 | → | 01 |
| 3 | 10 | 0 | 10 | 3 | 010 | ↓ | 10 |
| 4 | 11 | △ | 11 | 4 | 011 | ← | 11 |
| | | | | 5 | 100 | | |
| | | | | 6 | 101 | | |
| | | | | 7 | 110 | | |
| | | | | 8 | 111 | | |

根据表 5-5 约定图 5-16 编码结果见表 5-6。

表 5-6

| 轮廓号 | 灰度级 | 起始点 | | 方　向　序　列 | | |
|---|---|---|---|---|---|---|
| | | 行 | 列 | 第一个方向 | 第二个方向…… | |
| 00 | 00 | 000 | 000 | 01 | 01 | 0101…… |
| 01 | 01 | 001 | 001 | 01 | 10 | 1100 |
| 10 | 10 | 001 | 100 | 10 | 10 | 01 01 10 11 11 00 00 00 |
| 11 | 11 | 101 | 001 | 01 | 11 | |

## §5.6　分形编码

分形编码是目前公认的三种最有前途的编码方法之一(子波变换编码,模型法编码和分形编码)。分形编码是将分形理论应用于图像编码之中。Mandelbrot 创立了分形学,在他眼里,分形就是无穷之路。在 1967 年,Mandelbrot 研究了"英国的海岸线有多长?"这个数学家的难题。他认为:海岸线的长度,取决于所用的测量标尺的长度。标尺越短,可测出的海岸线的弯曲便增

多，测出的长度也就越长。海岸线长度的增长率，就是它的分维数。用分形的观点看，分形集具有任意小的比例细节，或者说具有精细的结构；分形集具有某种自相似的形式，可能是近似的自相似或统计的自相似，在大多数令人感兴趣的情形下，分形集是通过迭代方式产生的，在1985年，Barnsley正式提出了迭代函数系统(Iterated Function Systems，IFS)[5-3]理论。并将其用于给自然景物如云、海岸线、厥类植物建立比较逼真的分形模型上获得了巨大的成功。既然分形几何能非常逼真地产生许多自然图像，那么，反过来，能否用来对图像实行压缩呢？更确切地说，即给定一幅图像，如何寻找一个迭代函数系统，这个迭代函数系统的吸引子能够很好地逼近原图像？Barnsley领导的研究小组于1986年提出了著名的“拼贴定理”，从理论上解决了分形图像压缩这一道问题，为分形图像压缩编码奠定了理论基础。1987年Barnsley成立了迭代公司，分形图像压缩编码理论研究开始与实际相结合。

Mandelbrot在数学领域的主要贡献就是认识到通过递归定义的形状在自然界确实是普通存在的。Hutchinson经过更详细的研究后提出几何对象的全貌与局部，在仿射变换的意义下，具有自相似结构。Barnsley发展了Hutchinson的工作，并将其系统化和公式化，正式命名为迭代函数系统IFS，从而可以解析地构造分形。IFS目前已成为研究分形最成功、最普及的数学模型，在计算机图形学、生命科学、图像压缩编码等许多领域都取得了巨大的成功。如果我们建立了自然景物，那么我们就很容易地在计算机上实现它：从任意初始图像出发，经过简单的IFS迭代，直至收敛。反过来，对于一幅给定的图像，如何去求得它的IFS呢？Barnsley提出的拼贴定理提供了解决的方案。

要编码图像$X$，我们可以用$X$的有限个子图经过适当的压缩映射后去覆盖它。所用子图数目越少越好”而拼贴结果与$X$越接近越好：如果它们完全一致，则IFS的吸引子与$X$将相同；否则，IFS的吸引子只能以一定的精度逼近编码图像$X$。因为任何实际物理信号都可以利用欧氏空间上的紧集来抽象，所以利用IFS吸收子可以对任意信号逼近到任意精度，但没有保证一定能取得压缩效果。只是靠简单地引入大量的变换来提高逼近精度，而几乎没有用到吸引子精细结构性质。为了得到比较好的压缩效果，我们总希望用尽可能少的相似或仿射变换的结果去覆盖原图。

最初的分形图像编码是通过人机交互式的方式实现的。即使一名熟练的操作者，在图形工作站上压缩一幅典型的图像需要100小时，解码过程需30分钟左右。1989年分形编码取得突破性的进展。Barnsley的博士生Jacquin提出一种基于分块的全自动分形图像压缩编码方法，坚冰从此被打破，分形图像压缩编码的研究迅速在全球范围内展开。Jacquin提出的分形编码算法构成了后来绝大多数分形编码方案的基础。

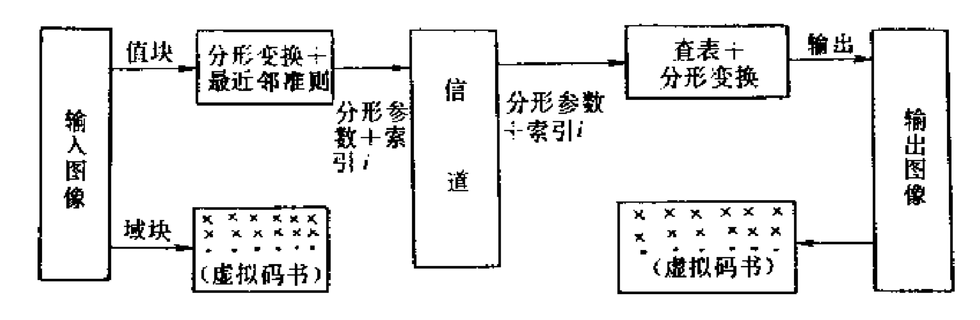



图5-17　分形编码原理框图-Jacquin算法

Jacquin 算法首先将编码图像划分成值块(range block)和域块(domain bloek)两类。其中,所有值块互不重叠,而他们的并集就是编码图像本身。域块之间可以重叠,尺寸比值块大一。其次,每个域块通过取平均的办法,得到与值块尺寸相同的图像块。所有尺寸缩小了域块的集合构成了虚拟码书。再次,对每一个值块,在虚拟码书中寻找距离最近的码字。在进行距离比较时,还对码字进行分形变换。其中包括对码字进行旋转,对折和镜像等几何变换并乘以一个合适的比例因子,作灰度平移操作。编码的结果除最佳码字的索引之外,还包括相应的分形变换参数。解码时,在虚拟码书中找到最佳码字,再对它实施相应的分形变换即完成一个图和块的解码过程。但分形解码是一个迭代过程,要进行多个输出图像到虚拟码书到输出图像的循环。

Jacquin 在提出全自动分形编码的同时,已经注意到了计算量大这个问题。为了降低计算量,他将图像块分成三类即:平坦块、一般块和边缘块。对不同类型采取不同的编码策略,这些措施降低了计算量,但效果不是很明显。

快速分形压缩编码沿着两条思路展开。第一通过合理组织域块,一般是以树结构的形式存放,实现域块的快速定位来减少编码计算量。利用这一技术,可以在编码图像质量损失不大的情况下将编码速度最大提高 50 倍。另一种思路是基于块分类的快速分形压缩编码研究,设图像总共有 $K$ 个域块,顺序搜索一遍需 $K$ 次块距离比较运算。如将其分发成 $m$ 大类,每大类包含几小类,这里 $K=m\times n$,此时要找到最佳匹配域块只需 $m+n$ 次块距离比较。明显地,对较大的 $m$ 和 $n$,$m+n$ 还小于 $m\times n$,因而能较大地提出编码速度。由于各类之间的几何变换关系可以事先通过人工确定,因而一旦知道值块和域块的分类结果就可以立即知道它们之间应实行的几何变换,避免了各种几何变换都需要尝试的缺点。还可以通过限定只有相同或相近类型的块才进行比较的办法进一步提高压缩速度。

自 Barnsley 提出分形图像压缩编码的概念以来,国内外在快速分形压缩编码方面进行了很多的探索。目前国内提出的基于 DCT 的块分类的快速分形编码方案是国内较好的一种算法,该算法在其损失 0.56~0.99dB 峰值信噪比的情况下,可以将传统的分形图像压缩速度提高 4.8~6.2 倍。图 5-19 是采用这种压缩方法对原图 5-18 的编码图像[5-4]。



图 5-18 Lena 原图(512×512×8bpp)



图 5-19 快速算法(33.54dB,13.60 倍,55 分钟)

另外基于 DCT 和自组织特征映射神经网络的块分类的分形编码算法的性能较基于 DCT 的块分类的分形编码等算法有较大的提高,可以达到国外典型的快速分形压缩编码的性能指标。分形—JPEG 混合编码是一种新的编码方法,它利用了 JPEG 和分形压缩编码的各自的优势,取长补短,使其性能较 JPEG 和常规分形压缩编码方法有明显提高。

总之,分形图像压缩编码是一门相当新的学科,其理论还在不断地发展和完善之中,分形

编码有其优越的一面，但要实用，还需要不断地探索。

## 习　题

5—1　试叙述哈夫曼编码方法和香农编码方法的理论依据，并扼要证明之。

5—2　试对题 5.2 图进行轮廓编码。

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 2 | 2 | 0 | 0 | 0 |
| 0 | 0 | 2 | 2 | 2 | 2 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 6 | 6 | 6 | 6 | 6 | 6 | 0 |
| 0 | 6 | 6 | 6 | 6 | 6 | 6 | 0 |
| 0 | 0 | 6 | 6 | 6 | 6 | 6 | 0 |

题 5.2 图

约定：（1）轮廓号、轮廓起始点位置和轮廓所包围区域的灰度级用二进制自然码。

（2）轮廓方向序列用二进制四向链码。

## 参 考 文 献


[5—1]　Pratt, W. K. et al Slant Transform Image Coding IEEE Trans, Comman COM-22, No. 8, 1974

[5—2]　王绍霖著. 数字图像处理. 长沙：国防科技大学出版社，1987

[5—3]　Barnsley M. F. Demko S. Iterated function systems and tle global construction of fractals, proceedings of the Royal Society of Landon A 399, 243～275 (1985)

[5—4]　邓元木. 分形在静止图像编码中的应用研究. 北京理工大学博士学位论文，1997 年 6 月

# 第六章　图像随机场模拟及处理

以上几章所介绍的图像处理方法基本上属于确定性的分析计算方法，而这一章对图像分析和处理的方法是统计的方法。在用统计方法对图像进行分析处理时，将对图像建立不同的统计模型，图像的演变亦遵循模型随机过程的统计规律，目前流行的一种模型是马尔可夫随机场模型(MRT)，这种模型是在1965年，由Abend Etal就开始研究，但直到70年代来，Hassher和Cross等人才真正利用马尔可夫模型以及其演生出的各种模型模拟出各种图像，特别是纹理图像，随后，Haluk Derin，S. Geman和D. Geman等人利用吉布斯模型(GRF)和高斯马尔可夫模型(GMRF)不断地把随机场模拟技术应用到图像的恢复、纹理分析以及边缘检测等各种领域，在图像处理领域亦取得了越来越显著的作用。我国的研究者，包括本书的作者和北京师范大学的匡锦瑜教授把以上随机模型视为随机神经网络的一种，并把广泛得到运用的模拟退火技术应用到图像恢复领域。同时对国外研究的技术作出改进，使低信噪比图像的恢复又增加了一种有力的武器。随着另一种优化方法—遗传算法的出现，遗传算法可以对复杂的非线性的多维数据空间进行迅速有效的计算，它的发展必将在图像的识别、恢复等各领域得到应用。本章将结合多年科研的积累，介绍用随机场模拟图像的原理与过程，介绍用吉布斯模型技术的随机神经网络模型、模型统计参数的估计、图像恢复中模拟退火技术以及其它技术的应用，最后将对遗传算法及其在图像处理中的应用作些介绍。

## §6.1　图像的随机场模型

### 一、图像的马尔可夫随机场模型

当我们探讨离散化图像的均值时，一般可用下式来表示

$$E\{X\}=[E\{X(n_x,n_y)\}] \tag{6.1.1}$$

式中$X$是图像阵列，$n_x$和$n_y$是图像的象点，$E$代表对$X$的数学期望。

人们还可以通过相关和协方差来研究数字图像中各点之间的关系。图像矩阵的相关函数为

$$R(n_1,n_2,n_3,n_4)=E\{X(n_1,n_2)X^*(n_3,n_4)\} \tag{6.1.2}$$

式中$(n_1,n_2)$和$(n_3,n_4)$表示两个不同位置的坐标点。同样地，图像矩阵协方差函数为

$$E(n_1,n_2,n_3,n_4)=E\{[X(n_1,n_2)-E\{X(n_1,n_2)\}]\times[X^*(n_3,n_4)-E\{X^*(n_3,n_4)\}]\} \tag{6.1.3}$$

可以直接写出图像矩阵的协方差函数为

$$\lambda^2(n_1,n_2)=k(n_1,n_2;n_1,n_2)$$

当图像$X$是一广义平稳过程，也就是说随机序列占据的整个$\pm\infty$空间内，相关情况只与

两采样点的间隔有关，而与采样点的位置无关，这样将有

$$R(n_1,n_2;n_3,n_4)=R(n_1-n_2;n_3-n_4)=R_{ij}$$

把这一概念引伸到协方差矩阵中去，协方差矩阵可以简写为

$$K_x=\begin{bmatrix} K_1 & K_2 & \cdots & K_{N_2} \\ K_2^* & & & K_{N_2-1} \\ \vdots & & & \vdots \\ K_{N_2}^* & K_{N_2-1}^* & \cdots & K_1 \end{bmatrix} \tag{6.1.4}$$

假定一幅图像用一个马尔可夫过程来表达，可以设法把以上协方差矩阵加以简化。

下面介绍马尔可夫链的概念：设 $X_1,X_2,X_3,\cdots,X_n$ 是一串随机变量，且有一阶马尔可夫链的性质，即条件概率

$$P(X_k/X_1X_2\cdots X_{k-1})=P(X_k/X_{k-1}) \tag{6.1.5}$$

对于所有 $k>j$ 来说

$$P(X_k/X_1X_2\cdots X_j)=P(X_k/X_j) \tag{6.1.6}$$

同样，对于 $k<n$ 来说

$$P(X_k/X_1\cdots X_{k-1},X_{k+1}\cdots X_n)=P(X_k/X_{k-1},X_{k+1})$$

这就是说，在一阶马尔可夫链中，任意点只与前后相邻两点有关。类似地，对 $n$ 阶马尔可夫链来说有

$$P(X_k/X_1X_2\cdots X_{k-1})=P(X_k/X_{k-r}\cdots X_{k-1})$$

和

$$P(X_k/X_1X_2\cdots X_n)=P(X_k/X_{k-1},X_{k-2},X_{k+2}\cdots X_{k+r}) \tag{6.1.7}$$

按照马尔可夫链的概念，如果在图像的每一行中，相邻象素间的相关系数为 $\rho_k$，且 $0\leqslant\rho_k\leqslant 1$，而自相关为 1，则协方差矩阵为

$$K_x=\begin{bmatrix} K_R(1,1)K_c & K_R(1,2)K_c & \cdots & K_R(1,N_2)K_c \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ K_R(N_{2,1})K_c & \cdots & \cdots & K_R(N_2,N_2)K_c \end{bmatrix} \tag{6.1.8}$$

式中 $K_c$ 为 $X$ 中每列的协方差矩阵，且 $|K_c|=N_1\times N_2$，$K_R(i,j)$ 为 $X$ 中 $i$、$j$ 行的协方差，另外，$|K_R|=N_2\times N_1$，对于上述特殊情况，上式可以写出

$$K_R=\sigma_R^2\begin{bmatrix} 1 & \rho_R & \rho_R^2 & \cdots & \rho_R^{(N_1-1)} \\ \rho_R & & & & \rho_R^{(N_1-2)} \\ \vdots & & & & \\ \rho_R^{(N_2-1)} & \cdots & \cdots & \cdots & 1 \end{bmatrix} \tag{6.1.9}$$

式中 $\sigma_R^2$ 为沿一行的象素的方差。这样一幅图就可以用一个参量 $K_R$ 来表达。

图像随机场模拟就是建立在马尔可夫过程的概念之上，一幅图像就可以视为二维的随机

过程，按照这样的概念形成随机图像，在随机场模型中，一幅图可以用一个联合概率密度来表示

$$P(X)=P\begin{bmatrix} x(1,1) & x(1,2) & \cdots & x(1,N_1) \\ \vdots & & & \\ \vdots & & & . \\ x(N_2,1) & \cdots & \cdots & x(N_2,N_1) \end{bmatrix} \tag{6.1.10}$$

上式说明可以用二维的联合概率分布来描述一幅随机场模型。如果用正态分布来描述此概率密度，则 $P(X)$ 将满足高斯分布的形式。

从以上的分析我们只能得到一系列的数学公式以及统计参数的公式，然而如何能够把深不可测的数学变成我们能够实际操作的方法，芬兰科学家 J. Besag 教授于 1974 年[6-1]的一个统计数学会议上谈到了"图像的统计分析与它的空间关系"，这篇文章哄动了整个科学界，由于它提出了解决以上难题的金钥匙，人们可以利用它提出的方法实现用计算机对图像的模拟。为此下面一一介绍几个必备的概念。

1. 邻域系统(Neighborhood System)

$L$ 的一个子集类 $u=\{u_{ij}:(i,j)\in L, u_{ij}\in L\}$ 称为 $L$ 的邻域系统，只要 $u_{ij}$ 是象素 $(i,j)$ 的邻域并满足条件：

(1) $(i,j)\notin u_{ij}$ 即 $(i,j)$ 点不包含在子集类中。

(2)对任意的 $(i,j)\in L$，若 $(k,e)\in u_{ij}$，则 $(i,j)\in u_{ke}$。为了进行图像模拟，需要利用以上定义的邻域系统 $u^1=\{u_{ij}^1\}$，$u^2=\{u_{ij}^2\}$，$u^m$ 称为 $m$ 阶的邻域系统，其各阶邻域系统如图所示。

| | | | 6 | | | |
|---|---|---|---|---|---|---|
| | | 4 | 3 | 4 | | |
| | 4 | 2 | 1 | 2 | 4 | |
| 6 | 3 | 1 | ij | 1 | 3 | 6 |
| | 4 | 2 | 1 | 2 | 4 | |
| | | 4 | 3 | 4 | | |
| | | | 6 | | | |

图 6-1　邻域系统

对于象素 $(i,j)$，它的 $u_{ij}$ 由四个相邻的象素点组成，$u_{ij}^2$ 由八个相邻的象素点组成。

2. 子团 $C$(Clique)

若 $L$ 的一个子集满足：(1) $C$ 由单个象素所组成，或；(2)对 $(i,j)\neq(k,e)$ 若 $(i,j)\in C$，$(k,e)\in C$，则 $(i,j)\in u_{ke}$，则称 $C$ 为 $(L,u)$ 对的子团。$(L,u)$ 的全部"子团"记为 $C=C(L,u)$。

有了以上几个定义，就可以介绍几种常用的随机场模型。

模型Ⅰ：马尔可夫随机场(Morkov Random Field)

(1) $P(X=x)>0$，对所有的 $X$ 其概率皆为正；

$$\begin{aligned}(2)\ &P[X_{ij}=x_{ij}/X_{ke}=x_{ke},(k,e)\neq(i,j)] \\ &=P[X_{ij}=x_{ij}/X_{ke}=x_{ke}\quad (k,e)\in u_{ij}]\end{aligned} \tag{6.1.11}$$

则称 $X$ 为以 $u$ 为邻域系统的马尔可夫随机场。注意这里的大写字母表示随机场或随机变量。小写字母表示一个具体的实现。从上式可以看到 MRF 是用条件分布表示的，亦就是用随机场的局部特征来表示。由于以上局部特征实现起来是比较困难的。所以马尔可夫随机场虽然已经研究了很久，但直到 70 年代末，80 年代初才开始蓬勃发展，其原因在于 F. Spitger 等科学家引入了 Gibbs 模型[6-2]，吉布斯模型起源于物理学与统计力学。科学家利用 Hammersley—Glifford 定理证明了 MRF 和 GRF 之间的一一对应的关系。为此可以设法由 GRF 来代替 MRF

实现随机场。GRF 避免了 MRF 的困难，由于它提供随机场的联合分布，而不是条件分布，为此它的实现就可以借助于实用的空间模型，亦即前面提到的邻域系统。因此 GRF 就广泛地应用到图像模拟的邻域之中。

模型 I：吉布斯随机场(GRF)

设 $M$ 为定义在 $L$ 上的邻域系统，若随机场 $X=\{X_{ij},(i,j)\in L\}$ 的联合分布具有以下的形式

$$P(X=x)=\frac{1}{Z}\mathrm{e}^{-v(x)} \qquad (6.1.12)$$

则称 $X$ 为吉布斯分布场。上式中 $V(x)=\sum_{C\in L}V_c(x)$ 为“能量函数”，其中 $V_c(x)$ 是子团 $C$ 的势，它仅与 $C$ 中 $x$ 的象素值有关，$Z$ 是归一化常数

$$Z=\sum_x \mathrm{e}^{-v(x)} \qquad (6.1.13)$$

$Z$ 称为配分常数。从定义式(6.1.12)可以看到。吉布斯分布是一种指数型分布，选择不同的能量函数 $V(x)$ 和子团势函数 $V_c(x)$，就可以构成不同形式的吉布斯分布模型。对于吉布斯的模型还可以作一般的物理解释，即若系统处于组态 $X$ 所具有的能量 $V(x)$ 越低，则系统处于该组态的概率就越大，即 $P(X=x)$ 就越大。

模型 II：高斯马尔可夫随机场(GMRF)

有关 GMRF 的定义式有多种。南加州大学著名教授 R. Chellappa 对 GMRF 有专著[6-3]专门研究了高斯马尔可夫随机场的各种定义、实现以及应用的实例，我们这里引用 Besag 所给的一种简单的定义[6-1]，如果 X 为高斯马尔可夫随机场 GMRF，则它的联合概率密度函数是

$$P(X)=(2\pi\sigma^2)^{-\frac{1}{2}n}|B|^{\frac{1}{2}}\exp^{\{-\frac{1}{2}\sigma^2(X-\mu)^T B(X-\mu)\}} \qquad (6.1.14)$$

其中 $\mu$ 是 $N\times1$ 的矩阵，$B$ 是 $n\times n$ 的方阵，它是正定对称的，它的对角线元素是统一的。

## 二、其它模型简介

### 1. 一维离散马尔可夫链[6-4]

(1)SSM(Striet－Sense Markov)　设$\{X_k,k\in k\}$是 SSM 马尔可夫链。如果对于任何 $n$ 值和 $K_1<K_2<\cdots<K_n$(在所有的 $K$ 值中)

$$P(X_{kn}=X_{ki},i=1,2,\cdots,n-1)=P(X_{kn}+X_{kn-1})$$

(注：以上 SSM 是有源的、单边的)

对于无源的的双边的情况有

$$P(X_{kg}=X_{ki},i=1,2,\cdots,n)=P(X_{kj},X_{k,j-1}\cdots X_{k,j+1}) \qquad (i\neq j)$$

对于一维离散的马尔可夫链，一维和二维的马尔可夫性质是相同的。

(2)WSM－N(WIDE－Sense Markov)(一维)

$$\hat{E}\{X_n|X_i,i<n\}=\hat{E}\{X_n|X_i,n-N\leqslant i<n\}$$

定理：一个 $W\cdot S\cdot S\{X_n\}$是 WSM－N(一维)，如果它有以下的性质

$$X_n = \sum_{e=1}^{N} heX_{n-e} + W_n$$

在这里：$\{he\}$是 LMMSEE 链的系数；$\{W_n\}$是估计的误差；$W_n \perp X_i$；对所有的 $i<n$；$M_m \perp W_n$（对 $m \neq n$），以上定理是在 1927 年被 Wood 所证明的，但它必须在高斯的假定之下。同时，如果$\{X_n\}$也是高斯的，WSM－M 指出，它也是 SSM－N。

(3)WSM－N(二维，Wide－Sense Markov)

$$\hat{E}\{X_n|X_i \quad i \neq n\} = \hat{E}\{X_n|X_i \quad 1 \leqslant |n-i| \leqslant N\}$$

定理：如果一个 $W.S.S\{X_n\}$是 WSM－N(二维)，然而它被表示为

$$X_n = \sum_{\substack{e=-N \\ e \neq 0}}^{N} heX_{n-e} + V_n$$

在这里，$\{h_n\}$是 LMMSEE 系数，并且$\{U_n\}$是估计的误差，$U_n \perp X_i$，对所有的 $i \neq n$；

$$E\{U_nU_n\} = \text{MSE}, \hat{X}_n = E\{U_n^2\} \triangleq \rho$$

$$E\{U_nU_m\} = \begin{cases} \rho & n=m \\ -h_{n-m^{\rho}} & 0<|n-m| \leqslant N \\ 0 & 其它 \end{cases}$$

$\{h_e'\}$定义为

$$\{h_e'\} = \begin{cases} 1 & e=0 \\ -h_e & 0<|e| \leqslant N \\ 0 & 其它 \end{cases}$$

$\{h_e'\}$是一个非负的有限序列，因而它对于$\{U_n\}$是自相关的。

## §6.2 图像模拟的实现

以上介绍了多种随机场模型，如 MRF 马尔可夫随机场模型、GRF 吉布斯随机场模型和 GMRF 高斯马尔可夫机场模型，如何实现它们，还需要借助于前面介绍的子团以及邻域系统等概念，由于马尔可夫随机场用条件概率来表示，所以它是不易具体实现的，而吉布斯随机场是用联合概率来表示的，为此它可以借助空间邻域系统来实现。

### 一、用邻域系统实现 GRF 图像

前面已经介绍了邻域系统，假设一个二阶的邻域系统如右图：

|   |   |   |   |   |
|---|---|---|---|---|
|   |   | 3 |   |   |
|   | 2 | 1 | 2 |   |
| 3 | 1 | * | 1 | 3 |
|   | 2 | 1 | 2 |   |
|   |   | 3 |   |   |

二阶邻域系统由中间的一个象点，以及周围的 8 个邻域点组成的二阶邻域系统，它们的子团如下(包括 1 和 2)：

$[*, \alpha]$

$[* \quad *, \beta_1]$ $\begin{bmatrix} * & \\ & ,\beta_2 \\ * & \end{bmatrix}$ $\begin{bmatrix} & * & \\ & & ,\beta_3 \\ * & & \end{bmatrix}$ $\begin{bmatrix} * & & \\ & & ,\beta_4 \\ & * & \end{bmatrix}$

$$\begin{bmatrix} * & * \\ & ,\gamma_1 \\ * & \end{bmatrix} \quad \begin{bmatrix} * & \\ & ,\gamma_2 \\ * & * \end{bmatrix} \quad \begin{bmatrix} * & * \\ & ,\gamma_3 \\ & * \end{bmatrix} \quad \begin{bmatrix} & * \\ & ,\gamma_4 \\ * & * \end{bmatrix}$$

$$\begin{bmatrix} * & * \\ & ,\xi \\ * & * \end{bmatrix}$$

其中 * 号表示子团中的象元，$\alpha$、$\beta$、$\gamma$、$\xi$ 表示对于各子团象素之间配以的参数，例如 $\alpha$ 就是对于中心点所配以的参数，$\beta_1 \sim \beta_4$ 是相对于中心点上下、左右、交叉两点之间配以的参数，$\gamma$ 相对于在中心点上下左右四个角上三象点之间配以的参数，$\xi$ 是四象点之间配以的参数。

从前面的吉布斯模型可以看出，吉布斯随机场是与网络元素的势能函数密切相关的，而子团的势能函数定义为

$$V_c(X)=\begin{cases} -Q & \text{如子团中所有象素的值相同} \\ Q & \text{其它} \end{cases} \tag{6.2.1}$$

其中 $Q$ 代表相应子团中的参数 $\alpha,\beta,\gamma,\xi$。

对于单个象点的子团，这子团的势能被定义如下

$$V_c(X)=\alpha_k \qquad \text{对 } X_{ij}=\gamma_k \qquad (\gamma_k=1,2,3,\cdots,M)$$

对于整个的随机场，我们假定随机场是均匀的，子团的势能仅仅依赖于子团的形式和象素的数值，而与子团在网络上的位置无关。

调整上面的 $\alpha_k$，可以调整每个图像域中象素的比例，调整其它的参数就可以改变每个域的大小和方向。

根据吉布斯随机场的模型式(6.1.12)，以及子团势能式(6.2.1)和所给的参数，将进行 GRF 的图像模拟，具体操作如下：

(1)在计算机上产生 $N_1 \times N_2$ 个随机数，组成初始的数字图像。

(2)对每个象素的概率数，进行初始灰度级的量化，例如把 0～0.25 的概率数计为灰度数 1，把概率数 0.25～0.5 计为灰度数 2 等等，继续下去，直到得到初始的黑白灰度值图像。

(3)由势能公式 $P(X=x)=\frac{1}{Z}E^{-V(x)}$，确定参数 $\alpha,\beta,\gamma,\xi$，计算子团势能，计算出每一象点新的概率，并量化到新的灰度等级。

(4)对于图像中的每个象点重复第三步。以新的灰度级代替旧的灰度级数，完成整幅图像，重复多遍。使图像稳定，过程结束。

下面列出 GRF 实际模拟图像的结果：以下所示图像，尺寸为 64×64，灰度级取为 4，迭代次数一般为 35，迭代次数的选择决定于图像的稳定程度，因为图像的形成视为一个随机过程，只有重复足够多的循环次数，才能得到一幅稳定的图像。当 $\alpha,\gamma,\xi$ 为固定常数，仅改变配分参数 $\beta$ 时，得到以下域图像和纹理图像。

1.域图像(如图 6－2 所示)

2.纹理图像(所图 6－3 所示)

从吉布斯模型中可以看出 $\beta$ 的变化可以改变图像的方向。当各个 $\beta$ 参数相近时形成域图像，当各 $\beta$ 参数变化较大时，就会形成各类纹理图像，以上结果仅仅是 $\beta$ 参数的变化，当各参数

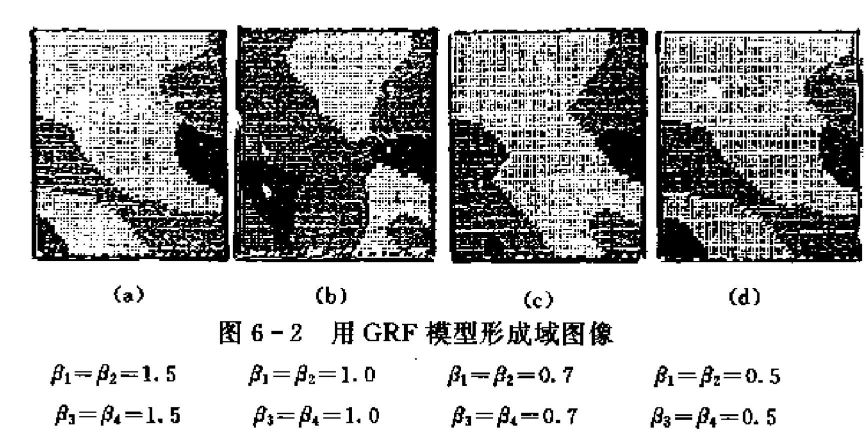


图 6-2 用 GRF 模型形成域图像

| (a) | (b) | (c) | (d) |
|---|---|---|---|
| $\beta_1=\beta_2=1.5$ | $\beta_1=\beta_2=1.0$ | $\beta_1=\beta_2=0.7$ | $\beta_1=\beta_2=0.5$ |
| $\beta_3=\beta_4=1.5$ | $\beta_3=\beta_4=1.0$ | $\beta_3=\beta_4=0.7$ | $\beta_3=\beta_4=0.5$ |

都发生不同变化时，将引出千姿百态的各类图像，形成一个庞大的图像库。

* * 以上吉布斯模型亦可以视为随机神经网络的一种。

前面已经介绍了吉布斯模型，GRF 与 MRF 可视为等价的，如 GRF 的条件概率

$$P(X_{ij}=x_{ij}\mid X_{k,e}=x_{k,e},\ (k,e)\neq(i,j))=\frac{P(X=x)}{\sum\limits_{X_{ij}}P(X=x)}$$

$$=\exp\left\{\frac{1}{Z}\sum_{(i,j)}\sum_{k}a_kV_{ij}^{(k)}(X_{ij},X_{ij}^{(k)})\right\}\Bigg/\sum_{(i,j)\in u_{ij}}\exp\left\{\frac{1}{Z}\sum_{(i,j)}\sum_{k}a_kV_{ij}^{k}(X_{ij}X_{ij}^{(k)})\right\} \tag{6.2.2}$$

在上式的分子、分母中，与 $X_{ij}$ 与 $X_{ij}^{(k)}$ 无关的势函数相同而被约去，因而式(6.2.2)仅与 $X_{ij}$ 及 $X_{ij}^{(k)}(k=1,2,\cdots,m)$ 有关，即式(6.2.2)是满足 MRF 的定义式(6.1.11)的。

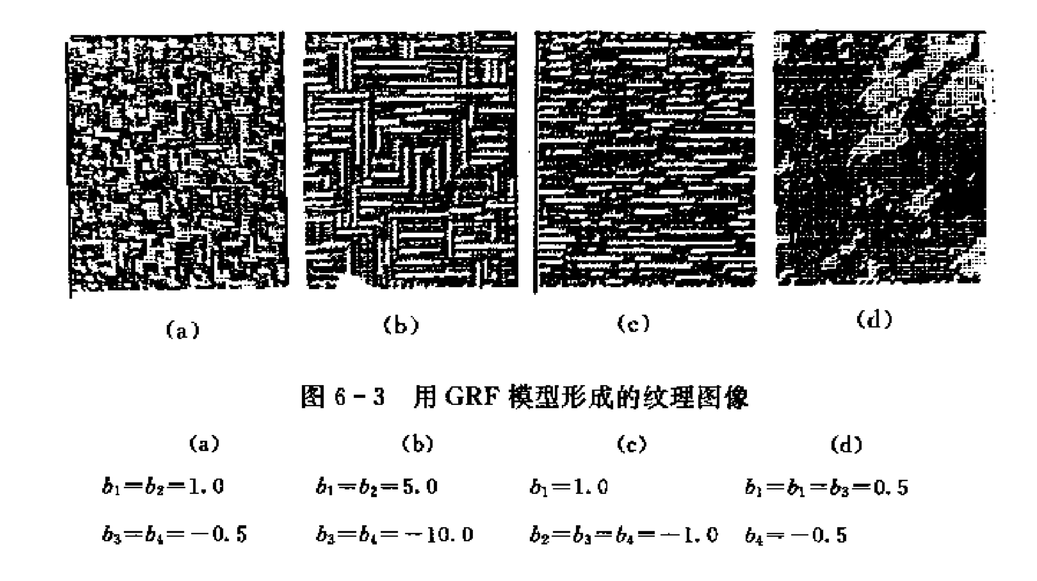


图 6-3 用 GRF 模型形成的纹理图像

| (a) | (b) | (c) | (d) |
|---|---|---|---|
| $b_1=b_2=1.0$ | $b_1=b_2=5.0$ | $b_1=1.0$ | $b_1=b_1=b_3=0.5$ |
| $b_3=b_4=-0.5$ | $b_3=b_4=-10.0$ | $b_2=b_3=b_4=-1.0$ | $b_4=-0.5$ |

当随机场随离散时间 $t,t+1,t+2,\cdots$ 而演化时，式(6.2.2)可以写成

$$P(X_{ij}(t+1)=X_{ij})=\frac{\exp\left\{\frac{1}{2}\sum_{(i,j)}\sum_{k}a_kV_{ij}^{(k)}(X_{ij},X_{ij}^{(k)}(t))\right\}}{\sum_{X_{ij}\in u_{ij}}\exp\left\{\frac{1}{2}\sum_{(i,j)}\sum_{k}a_kV_{ij}^{(k)}(X_{ij},X_{ij}^{(k)}(t))\right\}} \tag{6.2.3}$$

式中 $X_{ij}^{(k)}(t)$是 $t$ 时刻$(i,j)$的第 $k$ 个邻点的取值，$V_{ij}^{(k)}(X_{ij},X_{ij}^{(k)}(t))$是势函数。

上式网络的演化方法有多种，如概率演化法，最大似然法，以及 Metropolis 方法，我们仅介绍最后一种。

用 Metropolis 法进行系统模型的演化：从 $X_{i,j}$的当前状态 $X_{i,j}$出发，随机选择另一状态$X_{ij}'\in u_{i,j}$，根据式(6.2.2)计算概率比

$$R=\frac{P(X_{ij}(t-1)=X_{ij}')}{P(X_{i,j}(t-1)=X_{i,j})}=\exp\{-\Delta V(x)\} \tag{6.2.4}$$

式中$\Delta V(x)$为 $X_{i,j}(t-1)$分别取 $X_{i,j}'$,$X_{i,j}$时系统的能量差。当 $R>1$ 时，$X_{i,j}(t+1)$取 $X_{i,j}'$,当 $R<1$ 时 $X_{i,j}(t-1)$以概率 $R$ 取 $R'(i,j)$，以概率 $1-R$ 保留 $X_{i,j}$。当随机场的每个随机变量只取两种值 0,1 或－1,－1 时，则式(6.2.3)变成

$$P(X_{ij}(t-1)=X_{ij}')=\frac{1}{1-\exp\{-\Delta V(x)\}} \tag{6.2.5}$$

上式就是波尔兹曼模型，它是一种一般的神经网络模型。

## 二、用高斯马尔可夫随机场模拟图像(GMRF)

有关高斯马尔可夫随机场的内容美国教授 R. Chellappa 有专著，请看参阅文献[6-3]，它的定义方式亦是多种多样的，下面我们利用 Simultaneoas Autoregressive 方法或称同步自回归来实现图像的模拟。此方法比较简单，仅借用一阶邻域系统即可实现，任一象点 $X_{ij}$与其四周四个象点的关系为

$$X_{ij}=\beta_1' X_{i-1,j}+\beta_1X_{i+1,j}+\beta_2X_{i,j-1}+\beta_2' X_{i,j+1} \tag{6.2.6}$$

其中的 $\beta$ 参数说明了图像在水平和垂直方面上的变化，当 $\beta_1=\beta_1'=V_H$，$\beta_2=\beta_2'=V_R$ 时，上式变为

$$X_{ij}=V_H(X_{i-1,j}+X_{i+1,j})+V_R(X_{i,j-1}+X_{i,j+1})+\varepsilon_{i,j} \tag{6.2.7}$$

上式中的 $\varepsilon_{i,j}$就是均值为 0 方差为 1 的高斯随机变量，调正 $V_H$ 和 $V_R$ 就可以调正图像在二个方向的相对关系，一般取值小于 1，具体实现步骤如下：

(1)先编制一个子程序，实现一个点的高斯随机数，如果希望它的均值为 $\mu_y$，方差为 $\sigma_y^2$，则它是由原来的均值 $\mu_x$(可以为 0)和方差 $\sigma_x^2$(可以为 1)的高斯变量经过变换而来。其变换公式如下

$$Y=\left(\frac{X-\mu_x}{\sigma_x}\right)\sigma_y+\mu_y$$

其中 $Y$ 为 $N(\mu_y,\sigma_y^2)$，$X$ 为 $N(\mu_x,\sigma_x^2)$。

(2)重复调此子程序，使其布满幅图像，再利用前面的公式(6.2.3)，重新确定 $X_{ij}$点的象素值。以新的象素值代替旧的象素值，以完成整幅图像，其中 $V_R$、$V_H$ 是水平方向和垂直方向上可

变化的参数。

(3)式(6.2.3)中的 $\varepsilon_{ij}$是独立同分布的,均值为零,方差为1的白色噪声。

下面列出用高斯马尔可夫随机场模拟出的图像,在VAX11/780终端显示器上完成了高斯马尔可夫随机场模拟的过程,其象点数为128×128,这四幅图片的参数分别为

a)　$V_H=0.0$　　$V_R=0.0$

b)　$V_H=0.5$　　$V_R=0.5$

c)　$V_H=0.7$　　$V_R=0.7$

d)　$V_H=0.9$　　$V_R=0.1$

从上面可以看到利用GMRF模拟的图像多数近似为噪声图像,当水平垂直参数变化较大时,能出现水平和垂直条纹状的图像。

在以上图像模拟的过程中,有以下几点值得注意:

(1)以上的初始图像都是由计算机随机产生的,因此我们认为产生的图像仅仅是随机场的一个样本。

(2)用以上的GRF和GMRF可以很好地模拟出各种图像,需要我们进一步研究的是如何选择模拟参数以适应各种自然图像。另一课题是如何对一幅未知图像估计出它的参数,计算出它的统计特征参数,以便于图像的理解与识别,为此我们进行第三节参数估计技术的研究。



图6-4　图像用GMRF模拟的图像

## §6.3　图像参数估计方法的研究

前面两节介绍了用随机场模型及其计算机实现的方法。问题的另一方面,如果是一幅实际图像,如何估计出其参数,进而为其图像的进一步处理,如复原识别等等作准备。为此必须对统计参数进行估计。

### 一、混合高斯随机序列参数矩估计方法

这里提供的是对混合高斯随机数进行参数估计的一种新技术,使用一般的矩方程方法,首先建立样本矩和混合矩的方程。这些混合矩借助于各类别的参数来表示,而这些矩方程对未知数是非线性的,许多研究者已经进行了这方面的工作,但至今对此问题在一般情况下的完全解还不存在。作者在文章[6-4]中分析样本的各类别具有相同均值——标准差之比的情况下,使用Prony方法,可以把以上非线性方程转化为一组线性方程。这些方程不仅在理论上是正确的,而且对参数的估计是可行的,利用此改进后的矩方法可以对2、3、4类或更高类别的混合高斯随机数进行参数估计,经过实验得到了很好的结果。

计划在一幅模拟的数字图像中应用此方法，首先假定图像包含有 $C$ 个域，每个域有相同的灰度级 $u_i$，各个域的相对尺寸用 $P(i)$表示，$P(i)$表示 $i$ 类的数据占样本总容量之比。从而必然有$0\leqslant P(i)\leqslant 1$，$\sum_{i=1}^{C}P(i)=1$，图像中的每个象点都被模拟作为混合高斯随机数的一个样本。第 $i$ 类数据具有均值为 $\mu_i$，方差为 $\sigma_i^2$ 的高斯随机变量。我们的问题是：图像数据是由独立的 $N$ 个样本所组成，在类别数 $C$ 为已知的条件下，当各类参数 $\sigma_i/\mu_i=g$ 为常数的情况下，设法估计出每类随机数的参数 $\mu(i)$，$\sigma^2(i)$和 $P(i)$等等。

用一般矩方法解决以上问题，假设混合高斯数的样本容量为 $N$，混合后的概率密度函数是各个类别概率密度函数之和，即

$$f(x)=\sum_{i=1}^{c}P_if_i(x,\mu_i,\sigma_i^2) \tag{6.3.1}$$

这里 $C$ 是类别数，$f_i(x,\mu_i,\sigma_i^2)$是第 $i$ 类随机场数的概率密度函数。在以上四个参数 $C,P_i,\mu_i,\sigma_i^2$ 中，在 $\mu_i^2/\sigma_i^2=d$ 为常数并已知的情况下，$\mu_i$ 和 $\sigma_i^2$ 有一定的关系，所以只要知道 $\mu_i$ 就可以了。当 $C$ 为已知时，以上四参数中只需估计出 $\mu_i$ 和 $P(i)$即可。在 $d$ 未知的情况下，则必须首先估计出 $d$(或 $g$，$g=1/d$)值然后再估计出 $\mu_i$ 和 $P_i$ 之值。

对于一个高斯随机变量 $X$，它的 $K$ 阶原点矩可以表示为

$$m^{(k)}(P_1\cdots P_c;\mu_1\cdots\mu_c;\sigma_1^2\cdots\sigma_c^2)=\int_{-\infty}^{\infty}Z^kf(x)\mathrm{d}x \tag{6.3.2}$$

$f(x)$是混合后的概率密度函数

$$f(x)=\sum_{i=1}^{c}P_if_i(X,\mu_i,\sigma_i^2) \tag{6.3.3}$$

则 $K$ 阶混合矩可以表示为

$$M^{(k)}=m^{(k)}(P_1,P_2,\cdots,P_c,\mu_1,\mu_2,\cdots,\mu_c,\sigma_1^2,\sigma_2^2,\cdots,\sigma_c^2)=\sum_{i=1}^{c}P_im_i^{(k)}(\mu_i,\sigma_i^2) \tag{6.3.4}$$

式中 $m^{(k)}$是混合后的 $K$ 阶矩，$m_i^{(k)}$ 是第 $i$ 类数据的 $K$ 阶矩。式(6.3.4)是非线性的，待估计的参数在此方程之中，要直接进行运算是相当困难的，只有在类别数 $C=2$ 的情况下，经过一系列复杂的替代和变换，并在二类数据的方差相等的情况下才能进行运算并得到答案。如欲知细情，请参阅[6-5]，而在 $C>2$ 的情况下是不可做到的。

下面介绍用 Prony 方法，把以上非线性方程线性化，才能对多类数据进行分析运算。首先根据数学知识式(6.3.4)中的 $m_i^{(k)}$ 可以用 $\mu_i$ 和 $\sigma_i^{(k)}$ 多项式来表示。

$$m_i^{(k)}(\mu_i,\sigma_i^2)=\sum_{j=0}^{[K/2]}\gamma_{kj}\mu_i^{k-2j}\sigma_i^{2j}\qquad(k=1,2,\cdots,K;j=1,2,\cdots,C) \tag{6.3.5}$$

式中 $\gamma_{kj}$是可以查到的系数，它是 $X$ 的 $k$ 阶矩方程的系数 $\alpha,\beta$。

把式(6.3.5)代入式(6.3.2)中，得到矩方程为

$$M^{(k)}=\sum_{i=1}^{c}\sum_{j=0}^{[K/2]}P_i\gamma_{kj}\mu_i^{k-2j}\sigma_i^{2j}\qquad(k=1,2,\cdots,K) \tag{6.3.6}$$

该式仍是一个非线性方程，为了简化，假设各类数据满足 $d=\mu_i^2/\sigma_i^2$，设

$$g=\frac{1}{\sqrt{d}}\text{则}\qquad g=\frac{1}{\sqrt{d}}=\frac{\sigma_i}{\mu_i}\quad\text{或}\quad g^2=\sigma_i^2\mu_i^{-2}$$

则式(6.3.5)可以改写为

$$\begin{aligned}m_i^{(k)}(\mu,\sigma^2) &= \sum_{j=0}^{[K/2]}\gamma_{kj}\mu_i^{k-2j}\sigma_i^{2j}\qquad (i=1,2,\cdots,C)\\ &= \mu_i^{(k)}\sum_{j=0}^{[K/2]}\gamma_{kj}g^{2j}\\ &= \mu_i^{(k)}\alpha_k\end{aligned}$$

这里设

$$\alpha_k=\sum_{j=0}^{[K/2]}\gamma_{kj}g^{2j}\tag{6.3.7}$$

在 $g$ 为已知的情况下,$\alpha_k$ 也相应地确定了,对于式(6.3.4)可以写出

$$\begin{aligned}M^{(k)} &= \sum_{i=1}^{c}P_i m_i^{(k)}(\mu_i,\sigma_i^2)\\ &= \sum_{i=1}^{c}P_i\mu_i^k\sum_{j=0}^{[K/2]}\gamma_{kj}g^{2j}\\ &= \alpha_k\sum_{j=1}^{c}P_i\mu_i^k\end{aligned}\tag{6.3.8}$$

当 $g$(即 $d$)为已知时,$\alpha_k$ 也相应地确定,则式(6.3.8)可以写为

$$\sum_{i=1}^{c}P_i\mu_i^k=\frac{M^{(k)}}{\alpha^k}=Q_k(\alpha_k)\qquad (k=1,2,\cdots,K)\tag{6.3.9}$$

式(6.3.9)仍是一个非线性方程,现在我们再来考虑一个以 $\mu_i$ 值为根的多项式

$$(\mu-\mu_1)(\mu-\mu_2)\cdots(\mu-\mu_c)=0\tag{6.3.10}$$

将其展开后,可以得到一个 $\mu$ 的多项式

$$\mu^c-b_1\mu^{c-1}-b_2\mu^{c-2}\cdots b_{c-1}\mu-b_c=0\tag{6.3.11}$$

作如下的运算,把式(6.3.9)中的 $Q_{i+e}$逐项乘以 $b_{c-i}$,并把式(6.3.11)代入其中,并利用式(6.3.9)就可以得到如下公式

$$\sum_{i=0}^{c-1}Q_{i+e}(\alpha_k)b_{c-i}=Q_{c+e}(\alpha_k)\tag{6.3.12}$$

其中 $e=0,1,\cdots,K-C$

方程 (6.3.12)是一组以$\{Q_i\}_{i=1}$为未知数的线性方程 组,当 $g$ 为常数时,$\alpha_k$ 为常数,从式(6.3.9)不难求得样本矩 $M^{(k)}$,当 $Q_k(\alpha_k)$求得以后,用式(6.3.12)计算 $b_i$ 就不难了,这是一个线性方程,再将 $b_i$ 代入方程式(6.3.11)求得 $\mu_i$,再将 $\mu_i$ 代入方程式(6.3.9)得到 $P_i$,至此,全部参数估算完毕。从以上化简的过程可以看到,此方法是在一定的条件下,把 $2C$ 个非线性方程式(6.3.8)化为二组 $C$ 个线性方程式(6.3.12)和式(6.3.9)以及一个多项式(6.3.11),把此非线性计算转化为线性的计算。

在进行以上计算时注意两点:

(1)为使计算能够顺利进行,必须使 $K=2C-1$,可见类别数增加时,所需要计算的矩的阶次增加,计算变得复杂。

(2)以上分析是在 $g$ 为已知的情况下，式(6.3.12)是线性的，如果 $g$ 未知，可以首先把 $b_i$ 表示成 $Q_{c+e}(a_k)$与 $Q_{i+d}(a_k)$逆矩阵的乘积，再把 $b_i$ 代入式(6.3.12)的第 $C+1$ 个方程，此时 $e=C$，只有 $a_k$ 是未知数，从第 $C+1$ 方程求得 $g$(即 $a_k$)，再返回求 $b_i$，$\mu_i$ 和 $P_i$。在实际计算中，如果 $g$ 为未知，则要花费较大的精力。

## 二、实验结果及分析

用以上分析计算公式，在 VAX－750 和 IBM－PC 计算机上进行，类别数从 2，3，4 或更多，在容量 $N$ 以及参数 $\mu_i$，$P_i$，$g$ 不断变化的情况下得到下面的结果，下面四张图片(如图 6－5)分别反映出 $C=2$，$g=0.05$；$C=2$，$g=1$；$C=3$，$g=0.1$ 和 $C=4$，$g=0.1$ 四种情况下的的样本分布图。这些曲线是如此得到：首先按 $C$、$g$ 的不同值产生了所要求数量的随机数，然后把数据在 $XY$ 绘图仪，经过拟合形成样本分布的包络曲线，这四张图分别对应了相应实验的情况，在所给出的表 6－1 中，$A$ 是被说明的参数，它在模拟图像时被输入，$B$ 是在已知 $g$ 的情况下所估计出的参数，$C$ 是在未知 $g$ 的情况下，所估计出的数值。

表 6－1　高斯混合数的参数估计

| *Nu Des. g* | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | *Sam.* |
|---|---|---|---|---|---|---|---|---|---|
| 1. *A*. 0.1 | 80 | 160 | | | 0.5 | 0.5 | | | |
| *B*. 0.1 | 80.273 | 161.234 | | | 0.496 | 0.504 | | 2000 | |
| *C*. .098 | 80.343 | 160.180 | | | 0.499 | 0.501 | | | |
| 2. *A*. 1.0 | 80 | 160 | | | 0.6 | 0.4 | | | |
| *B*. 1.0 | 86.406 | 177.580 | | | 0.609 | 0.391 | | 2000 | |
| *C*. .946 | 77.330 | 166.669 | | | 0.583 | 0.418 | | | |
| 3. *A*. 0.1 | 80 | 160 | | | 0.5 | 0.5 | | | |
| *B*. 0.1 | 80.085 | 159.937 | | | 0.501 | 0.499 | | 5000 | |
| *C*. .102 | 80.063 | 159.784 | | | 0.501 | 0.499 | | | |
| 4. *A*. 0.01 | 80 | 160 | | | 0.8 | 0.2 | | | |
| *B*. 0.01 | 79.952 | 159.593 | | | 0.801 | 0.199 | | 100 | |
| *C*. .009 | 80.041 | 160.020 | | | 0.799 | 0.201 | | | |
| 5. *A*. 0.1 | 80 | 160 | 240 | | 0.3 | 0.3 | 0.4 | | |
| *B*. 0.2 | 80.741 | 163.816 | 240.066 | | 0.315 | 0.305 | 0.381 | | 2000 |
| *C*. .098 | 79.918 | 158.244 | 241.041 | | 0.299 | 0.298 | 0.402 | | |
| 6. *A*. 0.1 | 40 | 80 | 120 | 160 | 0.25 | 0.25 | 0.25 | 0.25 | |
| *B*. 0.1 | 40.226 | 80.967 | 125.320 | 162.570 | 0.218 | 0.264 | 0.267 | 0.251 | 2000 |
| *C*. .079 | 40.324 | 80.852 | 130.220 | 166.993 | 0174 | 0.277 | 0.292 | 0.257 | |

(1)比较实验 1、5、6 的结果可以看到，随着类别数 $C$ 的增加，估计精度下降。这是由于类别 $C$ 的增加则需要高阶的矩方程。为此增加了计算误差。

(2)比较实验 2 和 1 可以看到，随着 $g$ 的减小，均值与标准差之比增大，则估计的效果更好，图 6－5(a)和图 6－5(b)分别反映了 $C=2$ 而 $g=0.05$ 和 $g=1$ 的情况，后者两类数据非常靠近，为此增加了估计的难度。反之，$g$ 较小的情况，两类的数据分布远离，易于区别、估计效果好一些。这种现象同样适于多类别的情况。

(3)比较表中实验 1、3、4，随着样本数 $N$ 的增加，估计精度提高，可见样本数愈多，统计性能愈好。要注意在样本数为 100 的情况下，只有在 $C=2$ 和 $g$ 较小时可以得到满意的结果，当 $C$

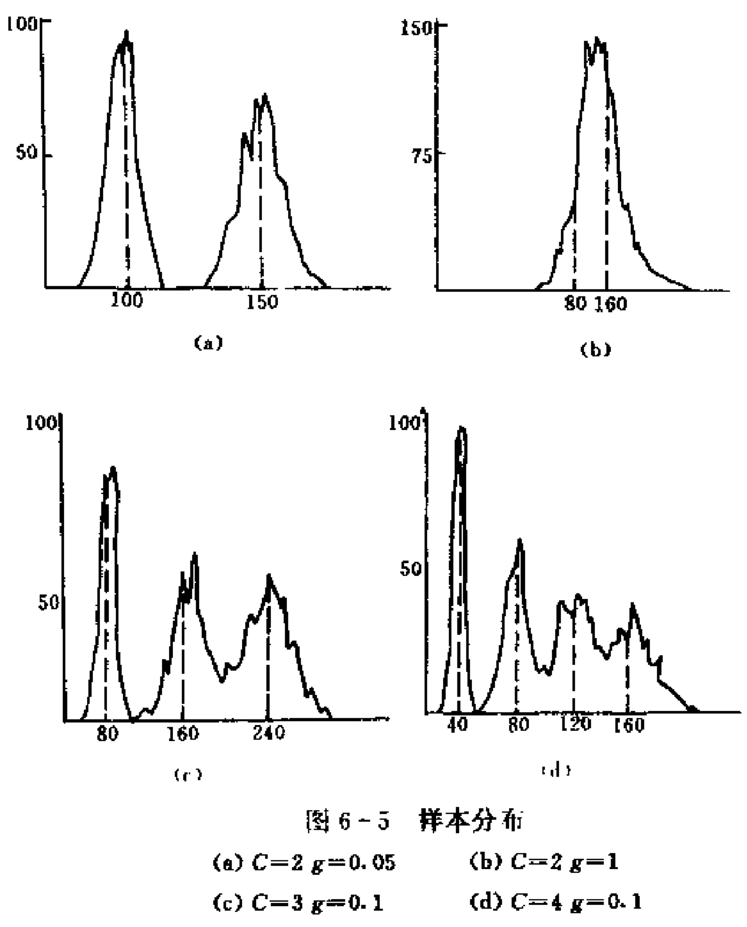



图 6-5 样本分布

(a) $C=2$ $g=0.05$ (b) $C=2$ $g=1$

(c) $C=3$ $g=0.1$ (d) $C=4$ $g=0.1$

较大，$g$ 较大时结果就没有如此理想，虽然增加 $N$ 值可以改善估计精度，但改善的程度并不大。

(4)在 $C=2$ 的情况下，一般估计的结果都很好，即使在样本容量 $N=100$ 非常小的情况下，估计的结果仍然很好，从实验 4 中可以看出。

(5)比较表中的实验说明，如果改变各类的比例 $P(i)$，对估计的精度影响不大。

综合以上各类可以看出，当 $g$ 为已知时估计的参数优于 $g$ 为未知时的参数，要想得到高质量的估计结果，需要增加样本容量，减少类别数 $C$，并减小 $g$ 的数值。与改进后的一般矩方法相比，计算简便并且可以用于 $C>2$ 的情况，对混合高斯随机数的参数估计将起很大的作用。

除矩估计法进行参数估计以外，还有其它多种参数估计方法，例如 E－M 方法，它是在最大化似然公式的过程中，建立起未知参数的循环递推公式，根据估计精度要求的高低将有不同的循环数目。

除了参数模型估计以外，进行图像复原的技术与前面讨论的图像模拟也有密切关系，有关模拟退火方法进行图像复原的技术已经在第四章中叙述，有关用 $E-M$ 算法及其在图像复原中的应用也可参考有关文献。

## §6.4 遗传算法及其应用

大多数生物体是通过自然选择和有性生殖这两种基本过程进行演化。自然选择的原则是适者生存，自然进化的这些特征早在60年代就引起了美国 Michigan 大学的 John Holland 的极大兴趣，他开始从事如何建立能学习的机器的研究。

Holland 注意到学习不仅可以通过单个生物体的适应而且通过一个种群的许多代的进化适应也能发生，受达尔文进化论的影响和适者生存的启发，他逐渐认识到，在机器学习的研究中，为获得一个好的算法，仅靠单个策略的建立和改进是不够的，还要依赖于一个包含许多候选策略的群体的繁殖，它起源于遗传进化，所以取名为遗传算法。

Holland 创建的遗传算法是一种概率搜索算法，它是利用某种编码技术作用于“染色体”的二进制数串，其基本思想是模拟由这些串组成的群体的进化过程。遗传算法通过有组织地又是随机地信息交换来重新结合那些适应性好的串，在每一代中，利用一代串结构中适应性好的位和段来生成一个新的串的群体；偶尔也要在串结构中尝试用新的位和段来替代原来的部分。遗传算法是一类随机算法，但它不是简单的随机走动，它可以有效地利用已有的信息来搜寻那些有希望改善解质量的串。类似于自然进化，遗传算法通过作用于染色体上的基因，寻找好的染色体来求解问题，与自然界相似，遗传算法对求解问题的本身一无所知，它所需要的仅是对算法所产生的每个染色体进行评价，并基于适应值来选择染色体，使适应性好的染色体比适应性差的染色体有更多的繁殖机会。

遗传算法利用简单的编码技术和繁殖机制来表现复杂的现象，从而解决非常困难的问题。特别是由于它不受搜索空间的限制性假设的约束，不必要求诸如连续性、导数存在和单峰等假设，以及其固有的并行性，遗传算法目前已经在最优化、机器学习和并行处理等领域得到了越来越广泛的应用。但需要说明，遗传算法类似于神经网络和模拟退火算法，这两种算法也是基于对自然界的有效类比，经过类比启示的开始阶段之后，遗传算法、神经网络以及 模拟退火算法已成为沿自身的道路发展下去的学科，它们距给它们启示的学科越来越远。

下面以一个非常简单的最优化问题为例来说明遗传算法。此例为四个连锁饭店寻找最好的经营决策，其中一个经营饭店的决策包括要做出以下三项决定：

- 价格　汉堡包的价格应该定在50美分还是1美元？
- 饮料　和汉堡包一起供应的应该是酒还是可乐？
- 服务速度　饭店应该提供慢的还是快的服务方式？

目的是找到个決定的组合(即经营决策)以产生最高的利润。

因为有三个决策变量，其中每个变量可以假设为两个可能值中的一个，所以对这个问题的每个可能的经营决策可以很自然地用长度 $e=3$，在规模 $k=2$ 的字母表上的特征串来表示。对每个决定变量，值0或1被指定为两个可能选择中的一个。这个问题的搜索空间包括 $2^3=8$ 个可能的经营决策。串长($e=3$)、字母表规模($k=2$)以及映射组成了对这个问题的表示方案，其中映射把串中具体位上的决定变量规定为0或1。利用遗传算法求解这个问题的第一步就是选取一个适当的表示方案。

按上面描述的表示方案，在8个可能的经营决策给出其中的4个。

| 饭店编号 | 价格 | 饮料 | 速度 | 二进制表示 |
|---|---|---|---|---|
| 1 | 高 | 可乐 | 快 | 011 |
| 2 | 高 | 酒 | 快 | 001 |
| 3 | 低 | 可乐 | 慢 | 110 |
| 4 | 高 | 可乐 | 慢 | 010 |

饭店的经营决策要由一位没有经验的新生决定从而他不知道在三个决定变量中哪个是最重要的，也不知道他做出最优决策下能得到的最大利润量或者他在做出错误决策可能招致的损失量，甚至不知道哪个变量的单独改变会产生利润上的最大变化。

新生不知道能否通过下面的逐步调整过程来接近全局最优值，在这个过程中每次改变一个变量，挑选好的结果，然后类似地改变另一个变量，再挑选好的结果。也就是说，他不知道变量能否单独地优化，或者它们是否以高度非线性方式相互联系。

新生面临另外的困难是只有通过每星期各个饭店的赢利情况来获得关于环境的信息。问题是他不清楚影响顾客光顾饭店的确切因素以及每个因素对顾客的决定起作用的程度。在营业过程中所观察到的饭店经营情况只是经营者从环境中得到的反馈，他不能保证经营环境在每个星期都保持不变，顾客的口味是多变的，并且决策的规则可能会突然改变，原来非常好的决策在某个新的环境中可能不再产生同样多的利润，环境的改变不仅可能是突然的，而且是不能预告的，通过观察到当前的经营决策不再产生与以前同样多的利润，经营者才会间接地发现环境的改变情况。

经营者还要面临的是要求立即做出经营决策，没有时间让他有单独的训练或单独的试验，唯一的试验来自实际营业的方式，此外，有用的决策过程必须立即开始产生一连串的中间决策，这些中间决策保持饭店从一开始到后续的每 个星期都在生存所需的最低水平之上。

因为经营者不了解他所面临的环境，他开始可能会明智地对四个饭店分别采用不同的初始随机决策，可以期望随机决策的获利近似地等于在搜索空间的总体上的平均获利。这样多样性一方面大大增加了获得接近于搜索空间内总体平均利润的机会，另一方面把从第一个星期的实际营业中学到的信息增加到最大限度。我们采用前面提供的 4 个不同的决策作为经营决策的初始随机群体。

事实上，饭店经营者是按与遗传算法同样的方式进行决策的。遗传算法的执行开始时是通过检测在搜索空间中随机选取的某些点来尽量学习关于环境的信息。特别地，遗传算法从第 0 代（初始随机代）开始，初始群体由随机产生的个体组织，在这个例子中，群体规模 $N$ 等于 4。

在遗传算法中，每一代群体中的个体都要在未知环境进行检测以得到它们的适应值，这里适应值取为利润，它还可以是获利、效用、目标函数值、得分或其它一些值。在这个问题中，初始群体的 4 个个体的适应值由表 6-2 给出。其中适应值简单地定义为，每个二进制染色体所代表的十进制值，所以决策 110 的适应值是 6 美元，全局最优适应值为 7 美元。

通过检测 4 个随机决策，经营者获悉到什么呢？表面上他知道了搜索空间中被检测的四个特殊点（即决策）的具体适应值（即利润）。特别地，他了解第 0 代群体中最好的个体 110 每周产生 6 美元的利润，最差的个体 001 每周只产生 1 美元的利润。

在遗传算法中用到的唯一信息是实际出现在群体中个体的适应值。通过模拟生物界自然选择和自然遗传过程，遗传算法把一个群体换到一个新的群体。一个简单的遗传算法由复制，

杂交和变异三个遗传算子组成。

**表 6-2　初始群体中经营决策的适应值**

| | | 第 0 代 |
|---|---|---|
| $i$ | 串 $X_i$ | 适应值 $f(X_i)$ |
| 1 | 011 | 3 |
| 2 | 001 | 1 |
| 3 | 110 | 6 |
| 4 | 010 | 2 |
| 总和 | | 12 |
| 最小值 | | 1 |
| 平均值 | | 3.00 |
| 最大值 | | 6 |

复制算子把当前群体中的个体按与适应值成比例的概率复制到新的群体中。在第 0 代，群体中个体适应值的总和为 12，因为最好的个体 110 的适应值为 6，所以群体的适应值归因于个体 110 的部分是 1/2。按照与适应值成比例的选择，我们期望串 110 将在新的群体中出现三次，因为遗传算法具有随机性，所以在新的群体中串 110 有可能会出现 3 次或 1 次。甚至以微小的可能性出现 4 次或根本不出现。群体的适应值归因于个体 011，010 和 001 的部分分别为 1/4，1/6 和 1/2，类似地我们期望个体 011 和 010 在新生的群体中分别出现 1 次，001 会从新的群体中消失。

复制算子的作用/效果是提高了群体的平均适应值，交配池群体的平均适应值是 4.25，而它起点的值仅为 3.00，而交配池中最差个体的适应值为 2，而在初始群体中最差个体的适应值为 1。因为低适应值趋向于被淘汰，而高适应值个体趋向于被复制，所以在复制运算中群体的这些改进具有代表性，但这是以损失群体的多样性为代价的。复制算子并没有产生新的个体，当然群体中最好的个体的适应值不会改进。

遗传杂交算子（有性重组）可以产生新的个体，从而检测搜索空间中新的点，复制算子每次仅作用在一个个体上，而杂交算子作用在从交配池中随机选取的两个个体上，杂交算子产生两个子代串，它们一般与其父代串不同，每个子代串都包含两个父代串的遗传物质。

杂交算子有多种，其中最简单的一点杂交算子的作用过程如下：首先产生一个在 1 到 $e-1$ 之间的一致随机数 $i$，然后配对两个串，相互对应地变换从 $i+1$ 到 $e$ 的位段。假设从交配池中选择编号为 1 和 2 的两个串为配对串，且杂交点选在 2（如下面的分隔符/所示），则杂交算子作用的结果为

01/1　　　010

11/0　　　111

一点杂交算子的重要特性是它可产生与原配对串完全不同的子代串，如上所示；另一个重要特性是它不会改变原配对串中相同的位，一个极端情况是当两个配对串相同时，杂交算子不起作用。

有充分的例子可以说明遗传算法利用复制和杂交算子可以产生具有更高平均适应值和更好个体的群体。

遗传算法从第 0 代到第 1 代，遗传算法迭代地进行以上过程，直到满足某个停止准则，在

每一代中，首先计算群体中每个个体的适应值，然后利用适应值的信息，遗传算法分别以概率 $P_r$、$P_c$ 和 $P_m$ 进行复制、杂交和变异操作，从而产生新的群体，停止准则有时表示成算法。

算法执行的最大代数目的形式，对那些一旦最优解出现就能识别的问题，算法可以当这样的个体找到时停止执行。

在这个例子中，第 1 代最好的经营决策 111 是：

· 汉堡包的价格定在 50 美分

· 饮料提供可乐

· 提供快速服务方式

经营决策 111 每周产生 7 美元的利润，是最优决策。如果我们恰好知道 7 美元是能够获的最大利润，那么在这个例子中可以在第 1 代就停止遗传算法执行。当遗传算法停止执行时，就把当前代中最好的个体指定为遗传算法的结果。当然，遗传算法一般不会像在这个简单例子中执行到第一代就停止，而是要进行到数十代、数百代、甚至更多代。

除了复制和杂交算子以外，变异算子也是遗传算法中经常用到的遗传算子，变异算子提供了一个恢复遗传多样性的损失的方法。

在准备应用遗传算法求解问题时，要完成以下四个主要的步骤：

(1)确定表示方案；

(2)确定适应值度量；

(3)确定控制算法的参数和变量；

(4)确定指定结果的方法和停止运行的准则。

在常规的遗传算法中，表示方案是把问题的搜索空间中每个可能的点表示为确定长度的特征串。表示方案的确定需要选择串长 $e$ 和字母规模 $k$，二进制串是遗传算法中常用的表示方法。在染色体串和问题的搜索空间中的点之间选择映射有时容易实现，有时又非常困难，选择一个便于遗传算法求解问题的表示方案经常需要对问题有深切的了解。

适应值度量为群体中每个可能的确定长度的特征串指定一个适应值，它经常是问题本身所具有的，适应值度量必须有能力计算搜索性空间中每个确定长度的特征串的适应值。

控制遗传算法的主要参数有群体规模 $N$ 和算法执行的最大代数目 $M$，次要参数有复制概率 $P_r$，杂交概率 $P_c$ 和变异概率 $P_m$ 等参数。

遗传算法的主要步骤如下：

(1)随机产生一个由确定长度的特征串组成的初始群体；

(2)对串群体迭代地执行下面的步①和②，直到满足停止准则；

①计算群体中每个个体的适应值；

②应用复制，杂交和变异算子产生下一代群体。

(3)把在任一代中出现的最好的个体串指定为遗传算法的执行结果。这个结果可以表示问题的一个解(或近似解)。

需要特别提到的是，遗传算法按不依赖于问题本身的方式作用在特征串群体上，遗传算法搜索可能的特征串空间以找到高适应值串，为了指导这个搜索，算法仅用到与搜索空间中检查过的点相联系的适应值，不管求解问题的本身，遗传算法通过执行同样的、惊人简单的复制、杂交和偶尔的变异操作来完成它的搜索。

在实际应用中，遗传算法能够快速有效地搜索复杂、高度非线性和多维空间。出人意外的

是遗传算法并不知道问题本身的任何信息，也不了解适应值度量。

北京理工大学电子工程系遗传算法研究小组，已经把GA算法运用到复杂图像的阈值分割，以及图像目标的快速识别方面，有关内容请参考[6-6]。

我们可以利用特殊领域的知识来选择表示方案和适应值变量，并且在选择群体规模、代数、控制执行各种遗传算子的参数、停止准则和指定结果的方法上也可以采取附加的判断，所有这些选择都可能影响到遗传算法在求解问题中的执行效果，甚至关系到它能否起作用。但总起来说，遗传算法仍是按不依赖于问题本身的方式快速搜索未知数的空间以找到高适应值的点。

# 习　题

6—1　请参阅文献，编制出初始随机场的初始图像，阵列为 64×64。其均值为 $\mu_y$，方差为 $\sigma_y^2$。

6—2　以题[6—1]得到的初始随机场为例，用讲义介绍的同步自回归方法形成高斯马尔可夫随机场，并改变其水平和垂直的参数，以观察其形成图像的结果。

6—3　以二阶邻域系统为基础，在题[6—1]初始随机场的基础上，选择图 6—3 的参数。形成域图像。灰度级取为 4。

6—4　以二阶邻域系统为基础，在题[6—1]初始随机场的基础上，选择图 6—4 所给的参数形成纹理图像。

# 参 考 文 献


[6—1]　besag J. Spatial Interaction and Statistical Analysis of Systems Cread before the Royal Statistical Society a meeting organiged by the research Section March 13th.

[6—2]　Spitger F. Markov Random Field and Gibbs Ensombles Ameriea Math. Mon. Vol 78

[6—3]　Chellappa R . chatterjee S . Classification of texture using Gaussian Markov Random Field IEEE Trans. Accoust Speech Signal Procees ASSP—33 1985 959—963

[6—4]　沈庭芝. 对混合高斯随机数参数矩估计方法的改进，电子学报，Vol. 19 No. 3 P110, 1991

[6—5]　Cohen A. C. Estimation in mixtures of two distributions Technometrics, Vo. 19 PP15—18 Feb. 1967

[6—6]　吴玲艳、沈庭芝、方子文. 遗传算法及其在图像分割及识别中的应用。硕士论文，1997

# 第七章 图像的分析

图像分析(也可称为图像理解)可以看作是一描述过程,主要研究用自动或半自动的装置和系统,从图像中提取有用的数据或信息,生成非图的描述或表示,这是当前图像处理与识别中的一个比较活跃的分支。

图像分析与增强、恢复、压缩编码等在处理上要求是不同的,它表现为系统的最终输出是数值、符号而不是图像。它也不同于经典的模式识别,即不仅限于给景物中诸区域在一定数目的已知类别内进行分类,还要给千变万化和难以预测的复杂景物加以描述。

图像分析的内容分为特征提取、图像分割、符号描述、纹理分析、运动图像分析和图像的检测与配准。

## §7.1 图像的特征

图像特征是指图像场的原始特性或属性。其中有些是视觉直接感受到的自然特征,如区域的亮度、边缘的轮廓、纹理或色彩。

有些是需要通过变换或测量才能得到的人为特征,如变换频谱、直方图、矩等等。

### 一、幅度特征

在所有的图像特征中最基本的是图像的幅度度量。可以在某一图像点或其邻区作出幅度的测量,例如在$(2N+1)\times(2N+1)$区域内的平均幅度,即

$$\bar{F}(i,j)=\frac{1}{(2N+1)^2}\sum_{m=-N}^{N}\sum_{n=-N}^{N}F(i+m,j+n) \qquad (7.1.1)$$

可以直接从图像象元的灰度等级,也可以从某些线性、非线性变换中构成新的图像幅度的空间来求得各式各样的图像的幅度特征图。

图像的幅度特征对于分离目标物的描述等都具有十分重要的作用。

### 二、直方图特征

我们知道,一幅数字图像可以看作是一个二维随机过程的一个样本,可以用联合概率分布来描述。通过测得的图像各象元的幅度值,可以设法估计出图像的概率分布,从而形成图像的直方图特征。

图像灰度的一阶概率分布定义为

$$P(b)=P\{F(i,j)=b\} \qquad (0\leqslant b\leqslant L-1)$$

式中$b$是量化层的值,共$L$层,$P(b)$是一阶近似直方图,它是

$$P(b)\approx\frac{N(b)}{M} \tag{7.1.2}$$

式中 $M$ 为围绕$(i,j)$点被测窗孔内的象元总数，$N(b)$是该窗口内灰度值为 $b$ 的象点数。

图像的直方图特征可以提供图像信息的许多特征，例如若直方图密集地分布在很窄的区域之内，说明图像的对比度很低，若直方图有两个峰值则说明存在着两种不同亮度的区域。

一阶直方图的特征参数有：

平均值 $$\bar{b}=\sum_{b=0}^{L-1}bP(b)$$

方差 $$\sigma_b^2=\sum_{b=0}^{L-1}(b-\bar{b})^2P(b)$$

歪斜度 $$b_s=\frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}(b-\bar{b})^3P(b)$$

峭度 $$b_k=\frac{1}{\sigma_b^4}\sum_{b=0}^{L-1}(b-\bar{b})^4P(b)-3$$

能量 $$b_N=\sum_{b=0}^{L-1}P(b)^2$$

熵 $$b_E=-\sum_{b=0}^{L-1}P(b)\log_2[P(b)] \tag{7.1.3}$$

二阶直方图特征是以象点对的联合概率分布的基础上得出的。若两个象元 $f(i,j)$及$f(m,n)$，它们分别位于$(i,j)$点及$(m,n)$点，两者的间距为$|i-m|$，$|j-n|$。并可用极坐标 $r$、$Q$ 表达，那么它们的幅度值的联合分布为

$$P(a,b)\triangleq P_k\{f(i,j)=a,f(m,n)=b\} \tag{7.1.4}$$

式中 $a$、$b$ 为量化的幅度值。为此，直方图估值的二阶分布为

$$P(a,b)\approx\frac{N(a,b)}{M} \tag{7.1.5}$$

式中 $N(a,b)$表示在图像中，在 $Q$ 方向上，经向间距为 $r$ 的象元对两点 $f(i,j)=a$，$f(m,n)=b$ 出现的频数，而 $M$ 是测量窗孔中象元总数。

假如图像的各象元对都是相互关联的，则 $P(a,b)$将在阵列的对角线上密集起来。以下列出一些度量，用来描述围绕 $P(a,b)$对角线能量扩散的情况：

自相关 $$B_A=\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}abP(a,b)$$

协方差 $$B_c=\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}(a-\bar{a})(b-\bar{b})P(a,b)$$

惯性矩 $$B_I=\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}(a-b)^2P(a,b)$$

绝对值 $$B_V=\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}|a-b|P(a,b)$$

能量 $$B_N=\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}[P(a,b)]^2$$

熵 $$B_E=-\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}P(a,b)\log_2[P(a,b)] \tag{7.1.6}$$

### 三、变换系数特征

由第二章的内容可知，二维变换得出的系数反映了二维变换后在频率域的分布情况。常可用二维傅里叶变换作为一种图像特征的提取方法。例如：

$$F(u,v)=\iint_{-\infty}^{\infty}f(x,y)e^{-j2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y \tag{7.1.7}$$

设 $M(u,v)$是 $F(u,v)$的平方值

$$M(u,v)=|F(u,v)|^2 \tag{7.1.8}$$

$M(u,v)$却与 $F(u,v)$不是唯一地对应的，即 $f(x,y)$的原点有了位移时，$M(u,v)$的值保持不变，$M$ 的这种性质称之为位移不变性，在某些应用中可利用这一特点。

如果把 $M(u,v)$在某些规定区域内的累计值求出，也可以把图像的某些特征突出起来，这些规定的区域如图 7-1 所示。

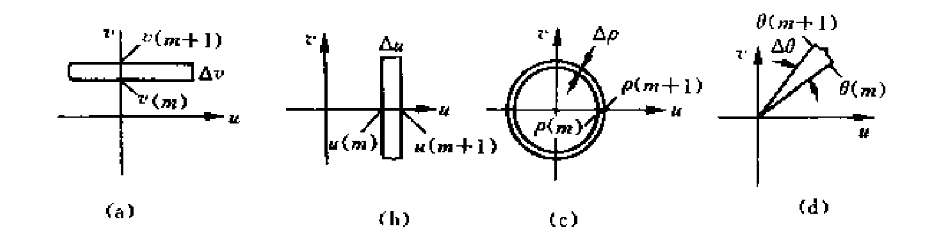



图 7-1 不同类型的切口

(a)水平切口； (b)垂直切口； (c)环； (d)扇区；

由各种不同切口规定的特征度量可由下式来定义，即

水平切口 $$S_1(m)=\int_{v(m)}^{v(m+1)}M(u,v)\mathrm{d}v$$

垂直切口 $$S_2(m)=\int_{u(m)}^{u(m+1)}M(u,v)\mathrm{d}u$$

环状切口 $$S_3(m)=\int_{\rho(m)}^{\rho(m+1)}M(\rho,\theta)\mathrm{d}\rho$$

扇状切口 $$S_4(m)=\int_{\theta(m)}^{\theta(m+1)}M(\rho,\theta)\mathrm{d}\theta$$

式中 $M(\rho,\theta)$是 $M(u,v)$的极坐标形式。

这些特征说明了图像中含有这些切口的频谱成分的含量。把这些特征提取出来以后，可以作为模式识别或分类系统的输入信息。如已经成功地运用到土地情况分类，放射照片病情诊断等各方面。

### 四、线条和角点的特征

图像中"点"的特征含义是，它的幅度与其邻区的幅度有显著的不同，检测这种点特征首先将图像进行低通滤波，然后把平滑后的每一个象元的幅度值与它相邻的 4 个象元的幅度值比较，当差值足够大时就可以检测出点特征来。

图像中线条的特征意味着它在截面上的幅度分布出现凹凸状，也就是说在线段的法向上，

图像的幅度是由低到高再到低(或相反)地变化的,可以用不同的掩膜来检测出线条。从图中提取这些特征,不仅可以设法压缩图像的信息量,也便于描述、推理和识别。

### 五、灰度边沿特征

图像的灰度、纹理的改变或不连续是图像的重要特征,它可以指示图像内各种物体的实际含量,图像幅度水平的局部不连续性称之为"边缘",大范围的不连续性被称之为"边界"。一个理想的边缘检测器应该能指出有边缘存在,而且还能定出斜坡中点的位置(精度达到一个象元)。

边缘检测的通常的方法是先对图像进行灰度边缘的增强处理,得出一个增强处理后的图像,然后设立门限,进行过门限操作来确定出明显边缘的象元位置。由于图像的空间幅度分布有时为正向,有时为负向变化,幅度值又具有慢变化的性质。所以门限值应随着空间总体幅度的变化而变化,门限设得过高,将漏掉小幅度变化的边缘,门限设得低,将出现由噪声引起的许多虚假的图像边缘。寻找一种对噪声不敏感、定位精确、不漏检真边沿又不引入假边沿的检测方法,始终是人们努力的目标。

### 六、纹理特征

通过对实际图片的观察,可以看到,由种子或草地之类构成的图片,表现的是自然纹理图像,由织物或砖墙等构成的图片,表现的是人工纹理图像。一般来说纹理图像中的灰度分布具有周期性,即使灰度变化是随机的,它也具有一定的统计特性。J. K 霍金斯认为纹理的标志有三要素:一是某种局部的序列性在该序列更大的区域内不断重复;二是序列是由基本部分非随机排列组成的;三是各部分大致都是均匀的统一体,纹理区域内任何地方都有大致相同的结构尺寸。当然,以上这些也只从感觉上看来是合理的,并不能得出定量的纹理测定。正因为如此,对纹理特征的研究方法也是多种多样的,有待于进一步探讨。

纹理可分为人工纹理和自然纹理。人工纹理是由自然背景上的符号排列组成,这些符号可以是线条、点、字母、数字等。自然纹理是具有重复性排列现象的自然景象。前者一般是有规则的,而后者往往是无规则的。

对纹理有两种看法,一是凭人们的直观印象,二是凭图像本身的结构。从直观印象出发包含了心理因素,这样就会产生多种不同的统计纹理特性,从这一观点出发,纹理分布应该用统计方法。如是从图像结构观点出发,则认为纹理是结构,纹理分析应该采用句法结构方法。描述纹理图像特性的参数有很多种,例如必须知道各个象素及其邻近象素的灰度分布情况。了解邻近象素灰度值变化的最简单方法是取一阶、二阶微分的平均值与方差。如要考虑纹理的方向性特性,则可考虑 $\Phi$ 方向与 $\Phi+\frac{\pi}{2}$ 方向差分的平均值与方差。

另一种方法是检查小区域内的灰度直方图。例如,取小区域为 $n\times n(n=3\sim7)$,作这 $n^2$ 个象素的灰度直方图。然后检查各小区域直方图的相似性,具有相似直方图的小区域同属于一个大区域,而直方图不同的小区域分属于不同的区域。

## §7.2 图像的分割

分割的目的是把图像空间分成一些有意义的区域。例如,一幅航空照片,可以分割成工业

区、住宅区、湖泊、森林等等。这里“有意义”内涵随着所需解决的问题的不同而不同，例如可以按幅度不同来分割各个区域，按边缘不同来划分各个区域，按形状来分割各个区域等等。

### 一、按幅度分割

幅度分割方法是把图像的灰度分成不同的等级，然后用设置灰度门限的方法确定有意义的区域或欲分割的物体之边界。

假定一幅图像具有如图 7－2 所示的直方图。由直方图(a)可以知道图像 $f(x,y)$ 的大部分象素取值较低，其余象素较均匀地分布在其它灰度级上。由此可以推断这幅图像是由有灰度级的物体叠加在一个暗背景上形成的。可以设一个阈值 $T$，把直方图分成两个部分，如图 7.2(b)所示。$T$ 的选择要本着如下的原则：$B_1$ 应尽可能包含与背景相关连的灰度级，而 $B_2$ 则应包含物体的所有灰度级。当扫描这幅图像时，从 $B_1$ 到 $B_2$ 之间的灰度变化就指示出有边界存在。当然，为了找出水平方向上和垂直方向上的边界，要进行两次扫描。也就是说，首先确定一个门限 $T$，然后执行下列步骤：

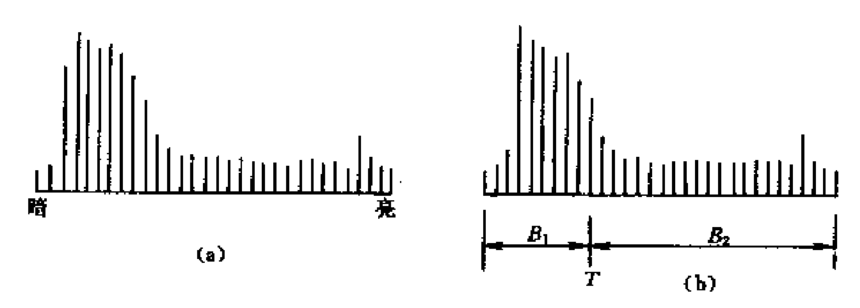



图 7－2　图像 $f(x,y)$ 的直方图

第一：对 $f(x,y)$ 的每行进行检测、产生的图像 $f_1(x,y)$ 的灰度将遵循如下规则

$$f_1(x,y)=\begin{cases}L_E & f(x,y)\text{和}f(x,y-1)\text{处在不同的灰度带上}\\ L_B & \text{其它}\end{cases} \tag{7.2.1}$$

式中 $L_E$ 是指定的边缘灰度级，$L_B$ 是背景灰度级。

第二：对 $f(x,y)$ 的每一列进行检测，产生的图像 $f_2(x,y)$ 的灰度将遵循下列规则

$$f_2(x,y)=\begin{cases}L_E & f(x,y)\text{和}f(x-1,y)\text{的灰度级处在不同的灰度带上}\\ L_B & \text{其它}\end{cases} \tag{7.2.2}$$

为了得到边缘图像，可采用下述关系

$$f(x,y)=\begin{cases}L_E & f_1(x,y)\text{或}f_2(x,y)\text{中的任何一个等于}L_E\\ L_B & \text{其它}\end{cases} \tag{7.2.3}$$

上述方法是以某象素到下一个象素间灰度的变化为基础的。这种方法也可以推广到多灰度级阈值方法中。由于确定了更多的灰度级阈值，可以提高边缘抽取技术的能力，其关键问题是如何选择阈值。

一种方法是把图像变成二值图像，如果图像 $f(x,y)$ 的灰度级范围 $(Z_1,Z_k)$，设 $T$ 是 $Z_1$ 和 $Z_k$ 之间的一个数，那么 $f_t(x,y)$ 可由下式表示

$$f_t(x,y)=\begin{cases}1 & \text{若}f(x,y)\geq T\\ 0 & \text{或}f(x,y)<T\end{cases}$$

另一种方法是把规定的灰度级范围变换为 1，而范围以外的灰度变换为 0，例如

$$f_u(x,y)=\begin{cases}1 & 若 f(x,y)\leqslant u\\ 0 & 或 f(x,y)>u\end{cases}$$

$$f_{u,v}(x,y)=\begin{cases}1 & 若 u\leqslant f(x,y)\leqslant v\\ 0 & 其余\end{cases} \tag{7.2.4}$$

另外，还有一种半阈值法，这种方法是将灰度级低于某一阈值的象素灰度变换为零，而其余的灰度级不变，仍保留原来的灰度值。总之设置灰度级阈值的方法不仅可以提取物体，也可以提取目标的轮廓。这些方法都是以图像直方图为基础去设置阈值的。显然，从直方图上妥善地选择 $T$ 值，对正确划分出感兴趣区域和背景很有关系。但哪里是最佳位置呢？这可用下述灰度值分布的概率模型来确定。

设图像中感兴趣的目标的象点灰度作正态分布，密度为 $P_1(x)$，均值和方差为 $\mu_1$ 和 $\sigma_1^2$，设背景点的灰度也作正态分布，密度为 $P_2(x)$，均值和方差为 $\mu_2$ 和 $\sigma_2^2$，换言之，整个密度函数可看作是两个单峰密度函数的混合。下面介绍最小错误分割法，设法找到一个门限，使划分目标和背景的错误分割概率为最小。

设目标的象点数占图像总点数的百分比为 $Q$，背景点占 $(1-Q)$，则混合概率密度为

$$\begin{aligned}P(X)&=QP_1(x)+(1-Q)P_2(x)\\&=\frac{Q}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right]+\frac{1-Q}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(X-\mu_2)^2}{2\sigma_2^2}\right]\end{aligned} \tag{7.2.5}$$

当选定门限为 $T$ 时，目标点错划为背景点的概率为

$$E_1(T)=\int_T^{\infty}P_1(x)\mathrm{d}X$$

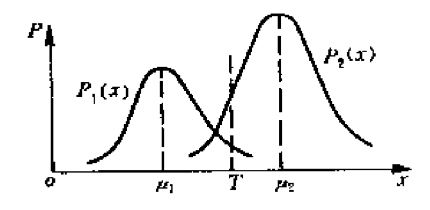


图 7-3　目标点和背景点的灰度分布

把背景点错划为目标点的概率为

$$E_2(T)=\int_{-\infty}^{T}P_2(x)\mathrm{d}X$$

则总的的错误概率为

$$E(T)=QE_1(T)+(1-Q)E_2(T)$$

令

$$\frac{\partial E(T)}{\partial T}=0$$

则有

$$-QP_1(T)+(1-Q)P_2(T)=0 \tag{7.2.6}$$

由此可得

$$\ln\frac{Q\sigma_2}{(1-Q)\sigma_1}-\frac{(T-\mu_1)^2}{2\sigma_1^2}=\frac{-(T-\mu_2)^2}{2\sigma_2^2}$$

当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时

$$T=\frac{\mu_1+\mu_2}{2}+\frac{\sigma^2}{\mu_2-\mu_1}\ln\frac{Q}{1-Q} \tag{7.2.7}$$

若先验概率为已知，例如 $Q=1/2$，则有

$$T=\frac{\mu_1+\mu_2}{2} \tag{7.2.8}$$

这表示正态分布时，最佳阈值可按式(7.2.7)、式(7.2.8)求得，若 $P_1(x)$，$P_2(x)$不是正态分布时，则可用式(7.2.6)确定最小误差的阈值 $T$。

对于复杂的图像，在许多情况下用单一的阈值不能给出良好的分割结果，在此种情况下，如果已知在图像上的位置函数描述不均匀照射可以设法用灰度级校正技术进行校正，然后采用单一阈值来分割，另外的方法是把图像分成小块，并对每一块设置阈值。但是，如果某块图像只含物体或只含背景，那么对这块图像就找不到阈值。这时，可以由附近的象块求得的局部阈值用内插法给此象块指定一个阈值。

在确定阈值时，如果阈值定得过高，偶然出现的物体点就会认作背景，如果阈值定得过低，则会发生相反的情况。克服的方法是使用两个阈值。例如 $t_1<t_2$，把灰度值超过 $t_2$ 的象素分类为核心物体点，而灰度值超过 $t_1$ 的象素仅当它们紧靠核心物体点时才算作物体点。$t_2$ 的选择要使每个物体有一些象素灰度级高于 $t_2$，而背景不含有这样的象素。同时，应选择 $t_1$ 使每个物体象素点具有高于 $t_1$ 的灰度级。如果只使用 $t_2$ 则物体总是分割得不完整，如果只使用 $t_1$ 则会有许多背景象素被错分为物体象素。如果同时使用 $t_1$ 和 $t_2$ 就能把背景和物体很好地分割开来。当然，如果物体与背景的对比是鲜明的，就不必使用这种方法。

此外，如果存在一个阈值 $t_2$，使得每个物体的象素灰度级高于 $t_2$，而背景不包含这种象素，可对图像设置阈值 $t_2$，然后检查高于阈值象素的邻域，目的是寻找一个局部阈值。以便在每个类似邻域中把物体和背景分开。如果这些物体相当小，并且不太靠近在一起时，这种方法比较适用。所使用的邻域应足够大，以保证它们既包含物体象素，也包含背景象素，这样就可以使邻域的直方图是双峰的。

有时需要寻找一幅图像的局部最大点，即提取比附近象素有较高的某种局部性质值的象素。一般来讲，也要求这些点具有高于一个低阈值 $t_1$ 的值，一旦超过 $t_1$，不管它的绝对值大小如何，一切相对的最大值都被采纳。因此，可以寻找局部最大值作为局部设置阈值的极端情况。在对图像进行匹配运算或检测界线时可采用这种方法。

## 二、边缘检测

作为图像的特征、图像的边缘和区域具有重要的意义，因此对边缘的检测和区域分割对于图像的分析和识别也是至关重要的。进行边缘检测的最基本的方法是图像的微分(即差分)，梯度和拉普拉斯算子等方法，有些基本内容已经在第三章中叙述过，这里主要介绍模板匹配，已知形状的曲线检测(Hoagh 变换)等等。

### 1. 模板匹配

在数字图像处理中，模板是为了检测某些不变区域特性而设计的阵列。模板可根据检测目的不同而分为点模板，线模板，梯度模板和正交模板等等。

点模板的例子如图 7-4 所示。下面用一幅具有恒定强度背景的图像来讨论。这幅图像包含了一些强度与背景不同且互相隔开的小块(点),假定小块之间的距离大于$[(\Delta x)^2+(\Delta y)^2]^{1/2}$,这里 $\Delta x,\Delta y$ 分别是在 $x$ 和 $y$ 方向的取样距离,用点样板的检测步骤如下:

| | | |
|---|---|---|
| −1 | −1 | −1 |
| −1 | 8 | −1 |
| −1 | −1 | −1 |

图 7-4 点模板

模板中心(标号为 8)沿着图像从一个象素移到另一个象素,在每一个位置上,把处在模板内的图像的每一点的值乘以模板相应方格中指标的数字,然后把结果相加。如果在模板区域内所有图像的象素有相同的值,则其和为零。另一方面,如果模板中心位于一个小块的点上,则其和不为零。如果小块在偏离模板中心的位置上,其和也不为零,但其响应幅度比起这个小块位于模板中心的情况要小一些,这时,可以采用阈值法清除这类较弱的响应,如果其幅度值超过阈值,就意味着小块检测出来了,如果低于阈值则忽略掉。

例如,设 $w_1w_2\cdots w_9$ 代表 3×3 模板的板,并使 $x_1x_2\cdots x_9$ 为模板内各象素的灰度值。从上述方法来看,应求两个矢量的积,即

$$W^TX = w_1x_1 + w_2x_2 + \cdots + w_9x_9 = \sum_{n=1}^{9} w_nx_n$$

$$W=\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_9 \end{bmatrix} \qquad X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix}$$

设置一阈值 $T$,如果

$$W^TX>T \tag{7.2.9}$$

我们认为小块已检测出来了。这个步骤可很容易地推广到 $n\times n$ 大小的模板,不过此时要处理 $n^2$ 维矢量。

线检测模板如图 7-5 所示,其中模板(a)沿一幅图像移动,它将对水平取向的线(一个象素宽度)有最强的响应。对于恒定背景,当线通过模板中间一行时出现最大响应,模板(b)对 45°方向的那些线具有最好响应,模板(c)对垂直线有最大响应,模板(d)则对−45°方向的那些线有最好的响应。

| | | |
|---|---|---|
| −1 | −1 | −1 |
| 2 | 2 | 2 |
| −1 | −1 | −1 |

(a)

| | | |
|---|---|---|
| −1 | −1 | 2 |
| −1 | 2 | −1 |
| 2 | −1 | −1 |

(b)

| | | |
|---|---|---|
| −1 | 2 | −1 |
| −1 | 2 | −1 |
| −1 | 2 | −1 |

(c)

| | | |
|---|---|---|
| 2 | −1 | −1 |
| −1 | 2 | −1 |
| −1 | −1 | 2 |

(d)

图 7-5 线模板

设 $w_1,w_2,w_3,w_4$ 是图 7-5 中四个模板的权值得出的九维矢量。与点模板的操作步骤一

样，在图像中的任一点上，线模板的各个响应为$W_i^T X$，这里$i=1,2,3,4$。此处$X$是模板面积内九个象素形成的矢量。给定一个特定的$X$，希望能确定在讨论问题的区域与四个线模板中的哪一个有最相近的匹配。如果第$i$个模板响应最大，则可以断定$X$和第$i$个模板最相近。换言之，如果对所有的$j$值，除$i=j$外，有

| a | b | c |
|---|---|---|
| d | e | f |
| g | h | i |

图 7-6　3×3 模板

$$W_i^T X > W_j^T X$$

则可以说$X$和第$i$个模板最相近。如果$W_1^T X > W_j^T X \quad j=2,3,4$，则可以断定$X$代表的区域有水平线的性质。

对于边缘检测来说同样也遵循以上原理，通常采用的方法是执行某种形式的二阶导数。类似于离散梯度计算，考虑 3×3 大小的模板，如图 7-6 所示。

在 3×3 的图像区域内、$G_x$及$G_y$分别用下式表示

$$\begin{aligned} G_x &= (g+2h+i)-(a+2b+c) \\ G_y &= (c+2f+i)-(a+2d+g) \end{aligned} \tag{7.2.10}$$

在$e$点的梯度为

$$G=[G_x^2+G_y^2]^{1/2}$$

采用绝对值的一种定义为

$$G=|G_x|+|G_y| \tag{7.2.11}$$

梯度模板如图 7-7 所示，把图 7-7 的区域与式(7.2.10)比较，可以看出$G_x$为第一行和第三行的差，其中最靠近$e$的元素($b$和$h$)的加权等于角偶上加权值的两倍，因此，$G_x$代表在$x$方向上导数的估值，式(7.2.11)可用图 7-7 中的两个模板来实现。

| 1 | 2 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| −1 | −2 | −1 |

| 1 | 0 | −1 |
|---|---|---|
| 2 | 0 | −2 |
| 1 | 0 | −1 |

图 7-7　梯度模板

2. 已知形状的曲线检测(Hough 变换)

当通过边沿检测得出一段小的边沿元时，人们常需知道它们是否能联成某一已知形状曲线，例如是否共直线、共圆等等，或者是图中是否存在某种形状的曲线？对这种问题 Hough 变换是一种有效的方法，它是把图像平面中的点按待求曲线的函数关系映射到参数空间，然后找出最大凝聚点，完成变换，下面介绍几种常用曲线的检测。

(1)寻找直线(线条检测[7-1])　极坐标系中直线的方程为

$$r=x\cos\theta+y\sin\theta$$

式中$r$是直线离原点的法线距离，$\theta$是该法线对$x$轴的角度，可见直线的 Hough 变换在极坐标系中是一个点，而一点的 Hough 变换是一正弦型曲线。计算 Hough 变换的方法是将$r-\theta$域量化成许多小格(见表 7-1)，对于每一个($x_0y_0$)点代入$\theta$的量化值，算出各个$r$，所得值(经量化)落在某个小格内，便使该小格的计数累加器加 1，当全部($x,y$)点变换后，对小格进行检验，有

大的计数值的小格对应于共线点，其$(r,\theta)$可用作直线似合参数。有小的计数值的各小格一般反映非共线点，丢弃不用。

可以看出，若$r$和$\theta$量化得过粗，则参数空间的凝聚效果差，找不出直线的准确的$r$、$\theta$值；反过来，若$r$、$\theta$量化得过细，则计算量将增大，因此需兼顾这两方面，取合适的量化值。

表 7-1

| $r$ \ $Q$ | 0° | 30° | 60°…150° |
|---|---|---|---|
| $r_1$ | | | |
| $r_2$ | | | |
| $r_3$ | | | |
| ⋮ | | | |
| $-r_1$ | | | |
| $-r_2$ | | | |
| ⋮ | | | |

若图像中各点是边沿元，而且梯度方向已求出，在寻找有无直线边沿时可在其梯度方向的一定范围内把$\theta$精细量化，其它$\theta$角则粗量化，这样在不增加总的量化小格数的情况下，可提高检测直线边沿的方向角的精度。

(2)寻找圆　对于圆，可写出其方程

$$(x-a)^2+(y-b)^2=R^2 \tag{7.2.12}$$

这时参数空间增加到三维，由$a$,$b$,$R$组成，如仍像找直线那样直接计算，计算量增大，不合适。若已知有圆的边沿元，而且边沿元为已知，则可降低一维处理，因为把上式对$x$取导数，有

$$2(x-a)+2(y-b)\cdot\frac{\mathrm{d}y}{\mathrm{d}x}=0 \tag{7.2.13}$$

这表示参数$a$和$b$不独立，利用式(7.2.13)以后，解关系式(7.2.12)只需用二个参数(例如$a$和$R$)组成参数空间，计算量缩减了很多。

(3)寻找椭圆：

在人为景物中圆形物体经常出现，透视成象后圆变成椭圆。为检测图中是否存在椭圆边沿，可仿照上述步骤进行。

设椭圆方程为

$$\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}=1$$

取导数有：

$$\frac{x-x_0}{a^2}+\frac{y-y_0}{b^2}\frac{dy}{dx}=0$$

可见这里有三个独立参数。如果椭圆主轴不平行于坐标轴，则可写为

$$Ax^2+Bxy+Cy^2+Dx+Ey+1=0$$

在利用椭圆边沿的方向信息后，在映射空间的独立参数仍将有四个之多，为了简化求椭圆的计算，有人研究了其他特殊解法，见参考文献[7-2]。

## 三、区域分割

对于特征不连续性的边缘检测，把图像分割成特征相同的互相不重叠连接区域的处理叫做区域分割。虽然目前已有许多方法[7-3]，但还都不是特别具有决定性的方法。因此有必要根据对象和目的不同而分别使用各种方法。前面所讲述的阈值处理，可以说是区域分割最简单的

方法。下面介绍其它几种方法。

1. 区域扩张法

(1)简单区域扩张法　作为区域分割的方法，最基本的有区域扩张法。这种方法一旦把图像分割成特征相同的小区域(最小的单位是象素)，就可以研究与其相邻的各个小区域之间的特征，把具有类似特征的小区域依次合并起来。例如，为了从象素开始进行区域扩张，可操作如下：

①对图像进行光栅扫描，求出不属于任何区域的象素。

②把这个象素的灰度与其周围的(4 邻域或 8 邻域)不属于任何一个区域的象素灰度相比较，如果其差值在某一阈值以下，就把它作为同一个区域加以合并。

③对于那些新合并的象素，反复进行②的操作。

④反复进行②、③的操作，直至区域不能再扩张为止。

⑤返回到①，寻找能成为新区域出发点的象素。

但是，用这样的方法，如图 7-8(a)那样的区域间边缘灰度变化很平缓的场合，或者如图 7-8(b)那样对比度弱的边缘相交为一点时，两个区域会合并起来。为了消除这一缺点，在②的操作中，不是比较区域外围象素的灰度与其周围象素的灰度，而是比较已经存在的区域的平均灰度与该区域邻接象素的灰度值。

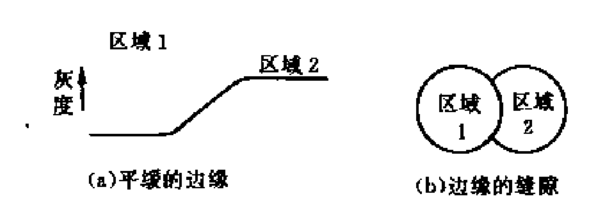



图 7-8　边缘对区域扩张的影响

但是，这样一来就会产生问题，从哪个象素起进行区域扩张，其最后的区域分割结果也就发生变化。

以下是不依赖于区域扩张起始点的方法。

①设灰度差的阈值=0，用上述的①～⑤进行区域扩张(使具有同一灰度象素合并)。

②求出所有邻接区域的平均灰度差，并合并具有最小灰度差的邻域区域组。

③通过反复进行②的操作，依次把区域合并。

用这种方法，如果在不适当的阶段停止区域合并，整个画面就会最终成为一个区域。

(2)统计假说检测法　以上的方法是把灰度差作为区域合并的判定标准的。此外，还有根据小区域内的灰度分布的相似性进行区域合并的方法。

①把图像分成相互稀疏的，大小为 $n\times n$ 的小矩形区域。

②比较邻接区域的灰度直方图，如果灰度分布情况都是相似的，就合并成一个区域。

③反复进行②的操作，直至区域合并完了为止。

为了检测灰度分布情况的相似性，采用下面的方法。这里，设 $h_1(X)$，$h_2(X)$为相邻的两个区域的灰度直方图，从这个直方图求出累积灰度直方图 $H_1(X)$和 $H_2(X)$，根据以下两个准则：

(1)Kolmogorov—Smirnov 检测

$$\max_X |H_1(X)-H_2(X)|$$

(2)Smoothed—Difference 检测

$$\sum_{X}|H_1(X)-H_2(X)|$$

求出两者之差，如果这个差在某一阈值以下，就把两个区域合并。这里灰度直方图 $h(X)$ 的累积灰度直方图 $H(X)$ 被定义为

$$H(X)=\int_0^X h(x)\mathrm{d}x \qquad \text{在数字场合 } H(X)=\sum_{i=0}^{X}h(i)$$

根据上述的灰度分布相似性的区域扩张法，不仅能为分割灰度相同区域使用，而且也能为分割具有纹理性的某个区域使用。但是，采用这种方法，如何把最初的 $n\times n$ 矩形区域作为单位，会出现下述情况：如果把 $n$ 定大了，则区域的形状就变得不自然，小的对象物就会漏过，相反，若把 $n$ 定小了，可靠性就会减弱。实际上，$n$ 常设在 5～10 的范围内。

(3)试探法　以上所有的方法都是采用了仅仅与灰度有关的值作为区域合并的标准。另外还有根据区域的形状作为判断标准的区域合并法。使用这种方法，首先把图像分割成灰度固定的区域，然后根据如下的评价函数，进行区域合并。

①把任意的邻接区域，$R_1$、$R_2$ 的周长设为 $P_1$、$P_2$，把在两个区域共同边界线两侧的灰度差在某一阈值 $\alpha$ 值以下的那部分长度设为 $W$。如果

$$W/\min\{P_1,P_2\}>\theta_1 \qquad \theta_1\text{:阈值}$$

则合并 $R_1$、$R_2$。

②在把 $R_1$，$R_2$ 的共同边界的长度设为 $B$ 的时候，如果

$$W/B>\theta_2 \qquad \theta_2\text{:阈值}$$

则合并 $R_1$、$R_2$。

①的标准是为了合并得到一致的理想形状的标准，$R_1$、$R_2$ 的共同边界在凸凹的情况下也易于被合并。而②则是为了合并共同边界中对比度低的部分比较多的区域的标准。

*2. 在特征空间利用群聚进行区域分割*

区域扩张法是重视图像空间的连通性进行区域分割的，与此相反，还有根据象素的相似性的区域分割法。

这种方法把图 7-9 所示的象素或小区域所具有的特征映射到特征空间中，根据在特征空间的群聚，求出具有相似特性的象素或小区域。以后，对各象素给予表示它所属的群的标号。为了在图像空间最后求得各区域，有必要对编有标号的图像进行连通成分的编号码操作。图 7-9 带有标号 1 的象素，被区分为两个连通区域。这种方法，是把象素或小区域作为 1 个图像的模式识别理论的应用。

对于图像内灰度或大小都不同的众多对象物存在的场合，简单的二值化方法起不到应有的作用。把简单的阈值处理加以扩充，可作为对复杂的图像能够顺利地进行区域分割的方法。这就是递归的阈值处理。这个方法是以彩色图像为对象而开发的，需进行如下处理。

①从彩色图像求出对应于红、绿、蓝、亮度、色调、彩度等一共为 9 种特性的直方图。

②从各个直方图求出峰，并选择最突出的峰。取出属于被选为峰的象素，并从这些象素里求出连通区域。

③对于用②求得连通区域(一般为多数个)以及其它剩余的连通区域，递归地反复进行①、②的处理，反复进行分割直至对于所有特征的直方图完全成为单峰性为止。这里所谓的递归就

是对于用①、②的操作得到的某一区域，再进行①、②的处理，并分割成若干个区域，再进一步对各个被细分的连通区域进行①、②的处理，如此逐次地把区域细分反复地进行下去。

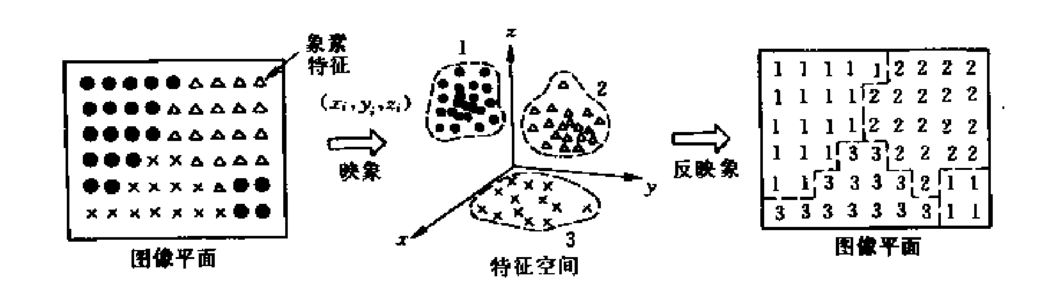



图 7－9　在特征空间中考虑群聚的区域分割

# §7.3　图像的纹理分析

在前面的边缘检测和区域分割的讨论中，是把灰度和颜色的一致性视为区域一致的。而纹理可认为是灰度和颜色的二维变化的图案，它是区域所具有的重要特征之一。不管什么样的物品（如白纸），如果一直放大下去再进行观察的话，就一定能显现出纹理（如纸的纤维花纹）。另外，也可以说，在灰度固定的区域中也是没有灰度变化的纹理。

纹理在图像处理中起到重要的作用。例如根据卫星摄影和航空摄影的地形和森林的分析，生物组织和细胞的显微镜照片的分析等等。此外，在一般的以自然风景为对象的图像分析中，纹理也具有重要的作用。

对纹理分析的方法，可分为统计方法和结构方法，统计方法常被用于象木纹、砂地、草坪那样的纹理而且不规则的物体，并根据关于象素间灰度的统计性质对纹理规定出特征。结构分析法，适合于象布料的印刷图案或砖花样等一类组成纹理的元素及其排列规则来描述纹理的结构。这里将介绍一下为求得纹理特征所需的各种统计的方法。结构方法也简单介绍之。

## 一、纹理特征及其计算

### 1. 直方图特征

直方图特征最基本的特征是纹理区域的灰度直方图或灰度的平均值和方差等等。由于用灰度的直方图，不能得到纹理的二维灰度变化，即使对于一般性的纹理识别，其能力也是过低的。如图 7－10 中两个纹理具有相同的直方图，只靠直方图就无法识别此图像，为此二维灰度变化的图案简单地赋予特征的方法，可以从图像求出边缘或灰度极大、极小点上的二维局部特征，并利用它们分布的统计性质的方法。即首先将图像进行微分从而求得边缘，作出关于边缘的大小和方向的直方图，并把这些直方图和灰度直方图合并，作为纹理特征。

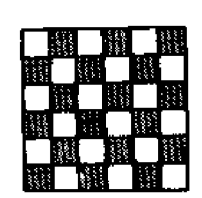
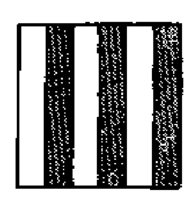



图 7－10　具有相同灰度的直方图纹理

另外，直方图较强的边缘密度或边缘的平均大小也成为表示纹理粗细的有用特征。如果关于边缘方向的直方图在某个范围内具有尖峰，那么就可以知道纹理所具有的对应于这个尖峰

的方向性。利用边缘的方向性，就可以较容易地识别图 7－10 的两种纹理。这样的特征，因为强烈地受到图像拍照时照明的影响，所以在进行特征计算前，有必要进行直方图的平坦化以及灰度平均值和方差的规格化。

2. 灰度共生矩阵特征

在灰度直方图中，因为各个象素的灰度是独立地进行处理的，所以不能很好地给纹理赋予特征。因此，如果研究图像中两个象素组合中灰度配置的情况，就能够很好地给纹理赋予特征。这样的特征，叫做二次统计量。其中有代表性的有以灰度共生矩阵为基础的纹理特征计算法。

灰度共生矩阵被定义为从灰度为 $i$ 的点离开某个固定位置关系 $\delta=(\mathrm{D}x,\mathrm{D}y)$ 的点上的灰度为 $j$ 的概率（图 7－11）。

$$P_\delta(i,j) \qquad (i,j=1,2,\cdots,n)$$

式中，$n$ 表示灰度级数而 $i$、$j$ 表示灰度，例如在图 7－11 的图像中，设 $\delta=(1,0)$ 的时候，$i=0$，$j=1$ 的组合（在 0 值的右邻为 1 的频率）有两次，即为 $P_{(1,0)}(0,1)=2$。另外在图 7－11 中，为表示 $\delta=(-1,0)$ 的关系而使用了相同的共生矩阵，用 $P_{(1,0)}(1,0)$ 表示其概率。因此，所有的共生矩阵 $P_\delta(i,j)$ 都是对称矩阵。

如果计算关于所有的 $\delta$ 的灰度共生矩阵，这就等于计算出了图像的所有二次统计量，但是，因为如果那样一来信息量就会过多，所以在实际中选择适当的 $\delta$，只对它求共生矩阵，多数场合使用图 7－11(c) 中所示的四种位移。作为纹理识别的特征量，不是原封不动地用上述的共生矩阵，而是要从各共生矩阵计算如下的特征量，并根据这些值给出纹理特征。

$$q_1=\sum_{i=1}^{n}\sum_{j=1}^{n}\{P_\delta(i,j)\}^2 \qquad q_2=\sum_{k=0}^{n-1}k\{\sum_{i=1}^{n}\sum_{j=1}^{n}P_\delta(i,j)\}_{|i-j|=k}$$

$$q_3=-\sum_{i=1}^{n}\sum_{j=1}^{n}P_\delta(i,j)\lg P_\delta(i,j)$$

$$q_4=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}i,jP_\delta P_\delta(i,j)-\mu_x\mu_y}{\sigma x\sigma y} \tag{7.3.1}$$

式中 $\mu_x=\sum_{i=1}^{n}i\sum_{j=1}^{n}P_\delta(i,j) \qquad \mu_y=\sum_{j=1}^{n}j\sum_{i=1}^{n}P_\delta(i,j)$

$$\sigma_x^2=\frac{1}{n}\sum_{i=1}^{n}(i-\mu_x)^2\cdot\sum_{j=1}^{n}P_\delta(i,j) \qquad \sigma_y^2=\frac{1}{n}\sum_{i=1}^{n}(j-\mu_y)^2\cdot\sum_{j=1}^{n}P_\delta(i,j)$$

除了这些特征之外，还有由共生矩阵计算的特征。根据一些实验，确认它的有效性，但是，因为所有的特征都是由数学上来定义的，所以究竟对应于哪一种纹理特征，对人来讲不太直观。另外，通常的图像灰度级 $n$ 一般要大到 256 左右，为了解决对特征计算费时间以及消除拍照时照明的影响，常常在求共生矩阵之前，根据直方图的平坦化预先就换成 $n=16$ 左右的图像。

3. 傅里叶特征

除了以上所述的图像空间提取特征之外，还有对图像傅里叶变换，从其频率成分的分布来求得纹理特征的方法。图像 $f(i,j)$ 的傅里叶变换 $F(u,v)$ 的功率谱，用下式定义：

$$P(u,v)=|F(u,v)|^2$$

其值表示了空间频率的强度。为了从 $P(u,v)$ 计算纹理特征，把它用极坐标的形式表示，并设为 $P(r,\theta)$ 之后，可求得

(a)移位 $\delta=(Dx,Dy)$

| 0 | 0 | 1 | 1 |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 2 | 2 | 2 |
| 2 | 2 | 3 | 3 |

(b)图像

$$P(1,0)=\begin{pmatrix}4&2&1&0\\2&4&0&0\\1&0&6&1\\0&0&1&2\end{pmatrix}\quad P(0,1)=\begin{pmatrix}6&0&2&0\\0&4&2&0\\2&2&2&2\\0&0&2&0\end{pmatrix}$$

0°方向　　90°方向

$$P(1,1)=\begin{pmatrix}2&1&3&0\\1&2&1&0\\3&1&0&2\\0&0&2&0\end{pmatrix}\quad P(1,-1)=\begin{pmatrix}4&1&0&0\\1&2&2&0\\0&2&4&1\\0&0&1&0\end{pmatrix}$$

135°方向　　45°方向

(c)灰度共生矩阵

图 7-11　灰度共生矩阵

$$P(r)=2\sum_{\theta=0}^{\pi}P(r,\theta)$$

$$q(\theta)=\sum_{r=0}^{W}P(r,\theta) \qquad (7.3.2)$$

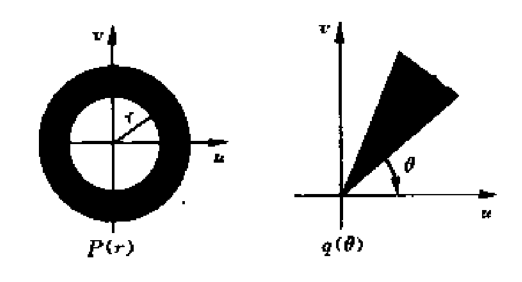



图 7-12　扇形滤波和环形滤波

如图 7-12 所示,$P(r)$是在功率谱空间以原点为中心的环形区域内的能量之和,$q(\theta)$表示扇形区域内的能量之和。作为纹理特征,使用 $P(r)$、$q(\theta)$图形的峰的位置和大小,$P(r)$、$q(\theta)$的平均值和方差等等。例如 $q(\theta)$的峰,表示纹理在其方向或直角的方向上具有明确的方向性,$P(r)$的峰,表示纹理构成元素的大小(纹理的粗糙程度)。

## 二、纹理区域的分割

上面所叙述的方法,是从具有同样的纹理特征的区域计算其特征的方法。在图像内存在着若干个不同的纹理区域的场合,为了利用这样的方法提取纹理区域,可以把图像分成 $n\times n$ 的小矩形区,在各矩形区内计算纹理特性。但是为了计算纹理特征,需要具有某种大小的小区域,所以用这种方法不能有效地产生细微的区域边界。

为了提取点密度不同的纹理区域,最好先计算以各点为中心的 $n\times n$ 区域内的点的密度,并求出密度直方图的峰。但是如果画面内存在多个结构区域的话,则不能用这种方法顺利地进行区域分割。为解决这一问题,在各点周围设置 5 个邻域,把在最一致邻域中的点密度作为该点输出值,再根据直方图进行分割,据此,应可以提取出细微的区域边界。

以上是把点密度作为纹理特征来使用。一般情况下,进行某种适当的滤波(例如把图像微分)之后,把以各点的中心局部区域中的边缘平均值的大小和方向作为该点的纹理特征,就能够应用上述的方法。

## 三、纹理边缘的检测

如同对于局部特征有边缘检测和区域分割两种方法一样,对于纹理特征,除了纹理区域的分割之外,还有纹理边缘的检测。如果用一般的边缘检测法,无法区别出依靠纹理区域内灰度变化图案所得的边缘和纹理区域之间的边缘。为了求得纹理边缘,可以分别求出$(i,j)$的 $n\times n$

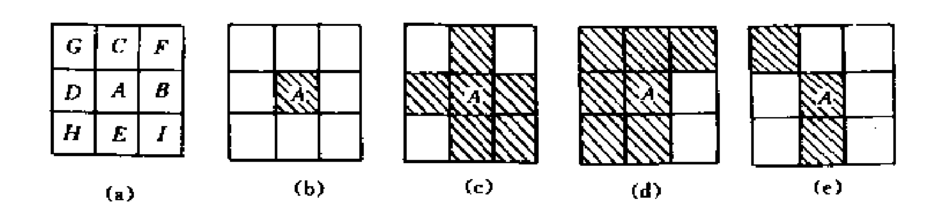



图 7－14　不同连通性质的象元

(a)象元位置；(b)孤立象元；(c)内部象元；(d)边界象元；(e)孤象元

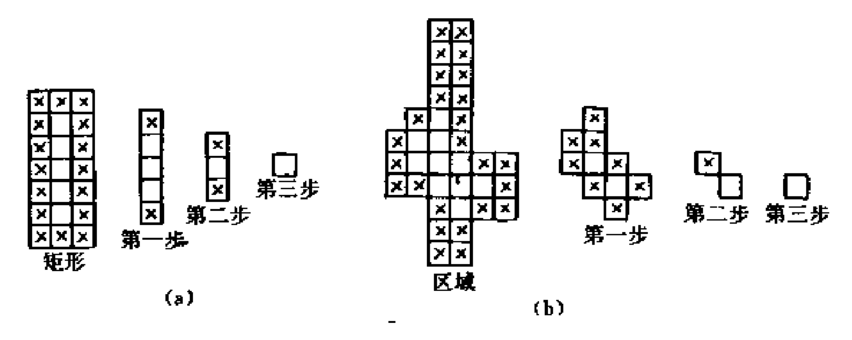



图 7－15　缩点操作举例

缩点操作是逐步把不是弧点的边界点（图中用×表示）移掉，直到只留下一点为止。在移掉边界点的过程中，不要导致引入由八方连通性质所定义的不连通区域。在不失去区域存在的条件下，可以移去弧端点。这种操作一直做到只留下一个点时为止。

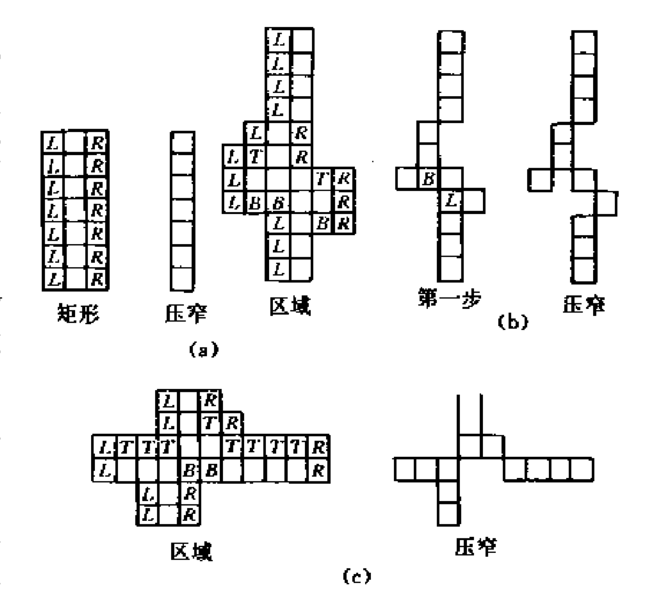



图 7－16　压缩操作举例

压窄的操作是先把物体的左面的边界点 $L$（不是弧线点）移掉，并保持物体的八方连通性质。然后再在同样的条件下把右面边界 $R$ 去掉，如图 7－16(a)所示，对于图 7－16(b)还可再移掉上部边界点 $T$ 及下部边界点 $B$。作了上述的移掉边界点的四步操作以后，还可在不影响连通性的条件下重复这种移掉边界的操作。显然若实施以上四个步骤的次序排列不同的话，将会出现不同的结果。

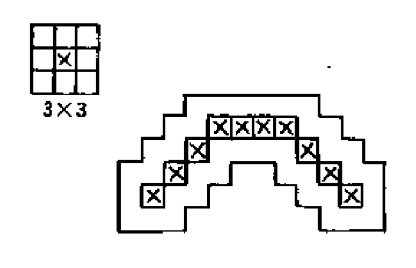



图 7－17　扩宽操作

扩宽操作是人为地扩大图像边缘的宽度。其操作方法是把每一个边缘点作为中心扩大成($n\times n$)的区域，$n$ 为奇数，区内的各象元的幅度与边缘点相同，如图 7－17 所示，这里 $n=3$。由图 7－17 看出用四小格构成的边缘上的每一个点，经(3×3)的八方扩宽操作以后，可得加粗的边缘如实线所示。

## 三、线条的描述，曲线拟合

若二维物体的边界或部分边界由一集点$(x_i,y_i)$，$i=1,2,$

$\cdots, M$ 所构成。其中的$(x_i, y_j)$是与$(x_{i+1}, y_{j+1})$相邻点。可以用某种函数关系

$$\bar{y}=g(x)$$

来拟合这种曲线见图 7-18,拟合的指标是使$(x_i, y_i)$与$(x_i, g(x_i))$之间的误差度量值最小。误差的典型度量值为

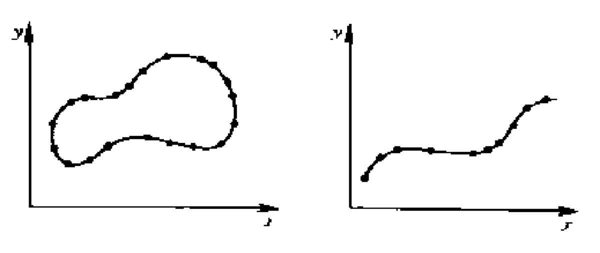


图 7-18 曲线拟合

绝对值误差和 $\varepsilon=\sum_{i=1}^{M}|y_1-g(x_i)|$

最小二乘方误差 $\varepsilon=\sum_{i=1}^{M}[y_1-g(x_i)]^2$

峰值误差 $$\varepsilon=\max|y_1-g(x_i)| \tag{7.4.1}$$

对于一般曲线的拟合来说,误差方程的最小化是很困难的。最一般的曲线拟合方法是用分段的多项式曲线来拟合。这时近似线段用下式来表达

$$\hat{y}=a_0+a_1x+a_2x^2+\cdots+a_Nx^N$$

把观测到的点的数据代进去,可得矢量空间关系式为

$$\begin{bmatrix} 1 & x_0 & x_0^2 & & x_0^N \\ 1 & x_1 & x_1^2 & \cdots & x_1^N \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_M & x_M^2 & \cdots & x_M^N \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \cdots \\ \hat{y}_M \end{bmatrix} \tag{7.4.2}$$

用矢量矩阵表达,得

$$Xa=\hat{y}$$

而最小二乘误差的准则是使

$$\varepsilon=(Y-\hat{Y})'(Y-\hat{Y})\rightarrow\min$$

多项式加权系数 $a$ 的最佳值可用广义逆矩阵来求得,即

$$a=x\hat{Y}$$

在数据点 $M$ 大于多项式系数 $N(M>N)$时,广义逆可用下式表达,即

$$X'=(X'X)^{-1}X'$$

这时可得 $$a=(X'X)^{-1}X'\hat{Y} \tag{7.4.3}$$

这里设定 $X_i$ 是单值的。可以看出这种求解方法与最佳线性估计方法一致的。

另外还有迭代式端点拟合,它是用分段的直线来拟合曲线,这里不多介绍。

### 四、形状描述

直线和曲线只是一般结构的基本单元,一般结构的例子有矩形,三角形,圆或其它各种闭合的形状。这种一般结构可以用它们的形状属性来描述分析。形状属性可分成:度量性的、拓扑性的、解析性的等多种。

1. 度量性的属性

一个图像度量性的属性是基于图像场内各点间的距离测量。两点$(x_i,y_i)(x_j,y_j)$之间的距离是一个实函数，它需要满足如下性质：

$$d[(x_i,y_i);(x_j,y_j)]\geqslant 0$$

$$d[(x_i,y_i);(x_j,y_j)]=d[(x_j,y_j);(x_i,y_i)]$$

$$d[(x_i,y_i);(x_j,y_j)]+d[(x_j,y_j);(x_k,y_k)]\geqslant d[(x_i,y_i);(x_k,y_k)]$$

满足以上定义性质的距离度量可以有以下几种：

(1)欧几里得距离

$$d_E=[(x_i-x_j)^2+(y_i-y_j)^2]^{1/2}$$

(2)幅度距离

$$d_M=|x_i-x_j|+|y_i-y_j|$$

(3)最大值距离

$$d_x=\max\{|x_i-x_j|,|y_i-y_j|\} \tag{7.4.4}$$

在离散图像中，后两种度量都可得整数，但欧几里得距离通常不是整数，因此会出现截断误差。度量性的属性还可用其它的形式来表达。例如最基本的一种是周长 $P$ 及面积 $A$。因此一种有用的属性就是窄度，它被定义为

$$T=4\pi\frac{A}{P^2}$$

例如圆形的 $T=1$，方形的 $T=\pi/4<1$；因此方形的窄度要比圆形的小。

形状的拓扑属性的描述，在没有撕裂或联接的情况下，这种形状不受任何变形的影响，并且不依赖于距离或由距离度量引出的任何性质。

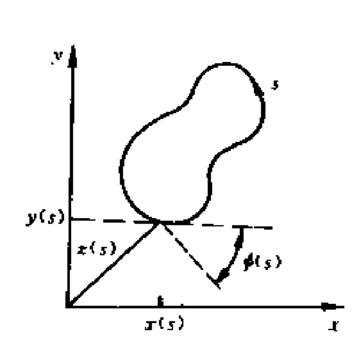



图 7-19 闭合曲线周长各点的切线

**2. 解析性的属性**

形状的解析属性是形状的一种数学描述。这种数学描述比用形状的位置及其离散样本值来描述形状要简单得多。解析性属性可用傅里叶变换来描述，还可以用矩的概念来描述。

(1)傅里叶描绘　一个任意闭合曲线，可以用它周长上的每一点的曲率来表示。若曲线周长上的一个点 $s$(或称为曲线上的弧长)在坐标系中的位置由下式定义，即

$$X(s)=x(s)+jy(s)$$

上式表示周长上弧长为 $s$ 的那一点的坐标为 $x(s)$、$y(s)$，由它们组成的矢量为 $X(s)$。在这一点上的曲线切向角可定义为

$$\mathrm{tg}[\varphi(s)]=\frac{\mathrm{d}y(s)}{\mathrm{d}x(s)}=\frac{\sin\varphi(s)\mathrm{d}\varphi(s)}{\cos\varphi(s)\mathrm{d}\varphi(s)} \tag{7.4.5}$$

根据曲线曲率的定义可知，弧长为 $s$ 的那点上的曲率 $k(s)$ 为

$$k(s)=\frac{d\varphi(s)}{ds}$$

故

$$d\varphi(s)=k(s)ds$$

而

$$\varphi(s)=\int k(s)ds$$

因为曲线形状是闭合的，所以随着周长上的弧长 $s$ 的增加，曲率 $k(s)$ 是一个周期性的函数。由于周期等于周长 $P$，所以基波频率为 $2\pi/P$。因此周期性函数 $k(s)$ 可以用傅里叶级数来描述，即

$$k(s)=\sum_{n=-\infty}^{\infty}C_n e^{jn(\frac{2\pi}{P})s}$$

而

$$C_n=\frac{1}{P}\int_0^P k(s)e^{-jn(\frac{2\pi}{P})s}ds \tag{7.4.6}$$

利用傅里叶系数 $C_n$ 来描述这条闭合曲线，这就叫做傅里叶描述。在实际应用中可以只选取 $(2m+1)$ 个傅里叶系数 $C_n(n=0,\pm1,\pm2,\cdots,\pm m)$ 来描述，这时采用的是截断了的傅里叶级数展开，可以利用它来逼近恢复原来的闭合形状。截断了的曲率 $k_m(s)$ 为

$$k_m(s)=\sum_{n=-m}^{m}C_n e^{jn(\frac{2\pi}{P})s} \tag{7.4.7}$$

利用下列关系式

$$dx(s)=\cos\varphi(s)d\varphi(s)=k(s)\cos\varphi(s)ds$$

$$dy(s)=\sin\varphi(s)d\varphi(s)=k(s)\sin\varphi(s)ds$$

可以利用 $km(s)$ 来恢复得出形状曲线上各点的坐标位置 $\hat{x}_m(s)$ 及 $\hat{y}_m(s)$，为

$$\hat{x}_m(s)=\hat{x}_m(0)+\int_0^s k_m(\alpha)\cos\varphi_m(\alpha)d\alpha$$

$$\hat{y}_m(s)=\hat{y}_m(0)+\int_0^s k_m(\alpha)\sin\varphi_m(\alpha)d\alpha \tag{7.4.8}$$

式中 $k_m(s)$ 已由前面给出，即

$$\varphi_m(s)=\int_0^s k_m(\alpha)d\alpha$$

而离散图像的情况下

$$X(s_j)=x(s_j)+jy(s_j)$$

$$\varphi(s_j)=\text{arctg}^{-1}\left[\frac{y(s_j)-y(s_{j-1})}{x(s_j)-x(s_{j-1})}\right] \tag{7.4.9}$$

而曲率由下式求得

$$k(s_j)=\varphi(s_j)-\varphi(s_{j-1}) \tag{7.4.10}$$

可以根据上式得出的离散的周期性曲率数据，来求得对应的离散傅里叶系数，使它来描述

这种闭合形状。物体形状的傅里叶描述具有旋转、变比、平移等不变性质。

(2)矩描述　假如给定一组边界点，可以使用傅里叶描绘子加以描绘。但有时一个区域内只有内部象元而又需要用一种随平移、旋转、变比而不变化的“描述子”来描述，那末，“矩”就是这样一种描述子。

若给定一个二维连续图像，它的灰度分布是 $f(x,y)$，则可以定义 $(p+q)$ 阶矩为

$$m_{pq}=\iint_{-\infty}^{\infty}x^p y^q f(x,y)\mathrm{d}x\mathrm{d}y$$

其中心矩为

$$m_{pq}=\iint_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q f(x,y)\mathrm{d}x\mathrm{d}y \qquad (7.4.11)$$

式中

$$\bar{x}=\frac{m_{10}}{m_{00}} \qquad \bar{y}=\frac{m_{01}}{m_{00}}$$

假如 $f(x,y)$ 是分段连续的，并在 $x-y$ 平面内只有有限部分是零值，则所有各阶矩都存在，并且矩序列 $(m_{pq})$ 唯一地决定了 $f(x,y)$。反之，$f(x,y)$ 也唯一地决定了 $(m_{pq})$ 序列。这种唯一性说明矩是一种代表图像 $f(x,y)$ 的一种特征。

对于离散图像 $F(j,k)$ 来说，矩和中心矩的公式如下：

矩
$$m_{pq}=\sum_j\sum_k j^p k^q F(j,k)$$

中心矩
$$\mu_{pq}=\sum_j\sum_k(i-\bar{j})^p(k-\bar{k})^q F(j,k) \qquad (7.4.12)$$

其中
$$\bar{j}=\frac{m_{10}}{m_{00}} \qquad \bar{k}=\frac{m_{01}}{m_{00}}$$

下面列出三阶以下的，用矩来表达的中心矩公式：

$$\begin{aligned}
&\mu_{00}=m_{00} && \mu_{11}=m_{11}-\bar{k}m_{10}\\
&\mu_{10}=0 && \mu_{30}=m_{30}-3\bar{j}m_{20}+2m_{10}\bar{j}^2\\
&\mu_{01}=0 && \mu_{12}=m_{12}-2\bar{k}m_{11}-jm_{02}+2\bar{k}^2m_{10}\\
&\mu_{20}=m_{20}-\bar{j}m_{10} && \mu_{21}=m_{21}-2\bar{j}m_{11}-\bar{k}m_{20}+2\bar{j}^2m_{01}\\
&\mu_{02}=m_{02}-\bar{k}m_{01} && \mu_{03}=m_{03}-3\bar{k}m_{02}+2\bar{k}^2m_{01}
\end{aligned} \qquad (7.4.13)$$

还可定义一种归一化的中心矩 $y_{pq}$，即

$$y_{pq}=\frac{\mu_{pq}}{\mu_{00}}$$

式中 $r=\frac{p+q}{2}$，而 $p+q=2,3\cdots$。从二阶及三阶矩可以导出一组 7 个不变矩，如下所示：

$I_1=y_{20}+y_{02}$

$I_2=(y_{20}-y_{02})^2+4y_{11}^2$

$I_3=(y_{30}-3y_{12})^2+(3y_{21}+y_{03})^2$

$I_4=(y_{30}+y_{12})^2+(y_{21}+y_{03})^2$

$$I_5=(y_{30}-3y_{12})(y_{30}+y_{12})[(y_{30}+y_{12})^2-3(y_{21}+y_{03})^2]+$$
$$(3y_{21}+y_{03})(y_{21}+y_{03})[3(y_{30}+y_{12})^2-(y_{21}-y_{02})^2]$$
$$I_6=(y_{20}-y_{02})[(y_{30}+y_{12})^2-(y_{21}+y_{12})^2]+4y_{11}(y_{30}+y_{12})(y_{21}+y_{03})$$
$$I_7=(3y_{12}-y_{30})(y_{30}+y_{12})[(y_{30}+y_{12})^2-3(y_{21}+y_{03})^2]+$$
$$(3y_{21}-y_{03})(y_{21}+y_{03})[3(y_{30}+y_{12})^2-(y_{21}+y_{03})^2] \tag{7.4.14}$$

这 7 个不变矩在连续图像条件下对平移、旋转、变比能保持不变的性质。在离散图像条件下，通过实验可得出在旋转 45°以下，比例放大 2 倍以下时仍具有保持不变的性质。

图像的这种矩法描述也是对图像的一种解析描述。

## 五、区域生长

区域生长是一种概念上的最简单的按幅度分割的方法，它把幅度相似的邻近的各个象元聚集在一起，构成一个分区。

科学家已发展并试验了一种区域生长的方法如下所述：先把幅度相似的并且具有上、下、左、右四个方向连通性质的量化象元对，组合起来构成一个"原子区"，然后把各"原子区"之间的弱边界"溶解"而使区域生长。在图 7-20 中有两个相邻区域 $R_1$ 及 $R_2$，它们的周长分别为 $P_1$ 及 $P_2$。这两个区域都是由许多子区域经溶解而合成的。$R_1$ 及 $R_2$ 之间的公共边界是 $C$。假如这个边界中有一段 $D$，在这一段边界的两边幅度小于某个值 $\varepsilon_1$，则这一段 $D$ 称之为"弱边界"。下面设定 $R_1$、$R_2$ 合并成一个区的法则，即

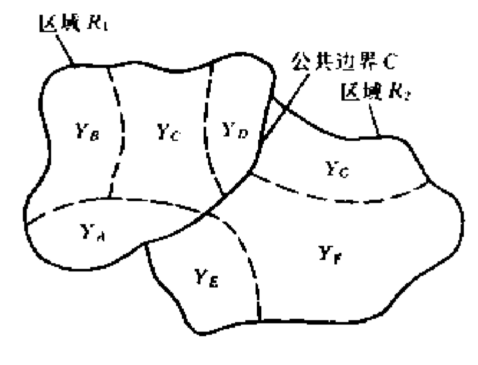



图 7-20　区域生长

$$D/\min\{P_1,P_2\}>\varepsilon_2$$

式中 $\varepsilon_2$ 一般取 1/2。这个法则意味着：若弱边缘 $D$ 的长度大于相邻区域 $R_1$、$R_2$ 中较小的一个周长[$\min\{P_1,P_2\}$]的一半以上，就可以使 $R_1$、$R_2$ 两个区合并成一个。这个法则促使小区域合并到大区域中去，而避免两个面积大小差不多的区域合并在一起。若使用了这个法则以后，弱边缘仍然保留，则还可用第二个法则，即

$$D/C>\varepsilon_3$$

一般取 $\varepsilon_3=3/4$，这意味着若弱边缘的长度占公共边界的 3/4 以上时，这可使相邻区域合并起来。假如只使用这后一个法则，将会使区域合并得太过份了。以上介绍的这种方法，在图像中景物比较简单、物体比较少、只有少量纹理的情况下比较适用，而不宜于在复杂的景物图像中使用。

## 六、形状分割

往往可以把一个复杂形状的物体分解成一组简单的、形状易于表达的物体。例如 T 字状物体可以分解成两个矩形来表达。这种划分的方法很简单，首先把任意形状的一个物体用一组连接起来的直线段或曲线段来拟合，然后可以在近似形状的各个转折点处进行分割。下面介绍多边形分割的法则为：把最邻近的凹点连接起来，构成各个半岛状的分区，然后把它们与主体分离开来，直到分区的所有部分都没有凹点时为止。图 7-22 示出了这个含义，有的科学家还

发展了这种多边形形状分区划分的算法。

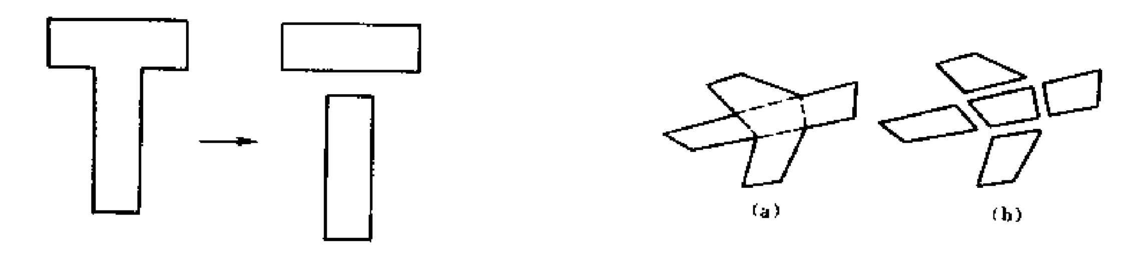



图 7-21 T字形状分割

图 7-22 多边形状分割

以上是关于区域形状的描绘,关于区域之间的关系描绘不再多写了。请读者参考其它的书籍。

## §7.5 多维信息及运动图像的分析和利用

前面的讨论,是把单幅图像作为处理对象来考虑的。随着图像处理的应用日益广泛,以各种各样的观点来观察图像,综合地分析由各种传感器测定的多幅图像数据,以期望提取精度更高和更多方面的信息的这类研究,正在积极进行着。多重图像分析的代表便是遥感,这部分内容将在第八章介绍。

除了上面所提到的多维信息以外,时间也是重要的特征。关于图像随时间变化的信息,可以根据对多幅图像的集合进行处理来规定其特征。在时间信息的场合,这种图像集合中的各幅图像是对应于把同一目标从时间上错开进行观测所得出的图像。利用遥感进行农作物和森林分类时,因为只根据光谱信息和纹理信息,因此有时实现不了细节和精度高的分析。在这样的场合,可以把相同的地区放在适当的时间进行拍照,把目标的光谱特性随时间变化作为用来进行分类的特征量加以利用。这表现在即使某个时期表示相同光谱特性的植物,如果随生长的速度和季节的变化而不同,也能够通过研究光谱特性的时间变化比较容易地加以识别。

常常把时间间隔比较短的序列图像叫做运动图像。在运动图像的分析中,图像内目标的变化极小,把这样小的变化一个一个地联系起来,提取出目标的运动信息和形变,三维形状,这就是分析的目的(图 7-23)。

由在时间上相邻的两幅图像求出目标的位置,形状变化的最简单的方法是取两幅图像的差。如图 7-24,当背景亮的目标物体移动时,在两幅图像的差分图像中,就形成了在移动的前方为正值的区域,在后方为负值的区域。用这种方法只能检测出运动物体的一部分,但可以推断来自正负区域组的移动矢量。为了从一系列的差分图像求出运动物体的形状,只要求出连续的差分图像中正的或负的区域的逻辑和就可以了(图 7-25)。

一旦运动物体被提取出来,就可以在以后的画面上跟踪它,从而求出时间的变化。此时,因为已经知道了运动物体的特征和移动速度,就可以利用这些信息重新在分析的图像里推断物体位置并有效地进行处理。

另外,一般从拍摄的物体重叠图像中把物体分离开来是比较困难的。这时,如果利用该图像中物体分离之前或之后的物体位置和移动矢量,就可以比较简单地分离重叠物体(图 7-26)。

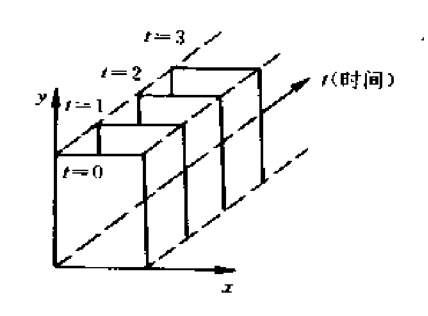


图 7－23　运动图像(时间序列图像)

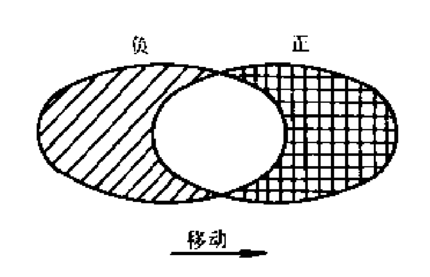


图 7－24　差分图像

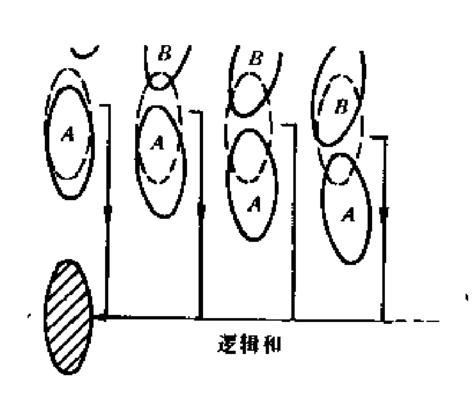


图 7－25　运动物体的形状的提取

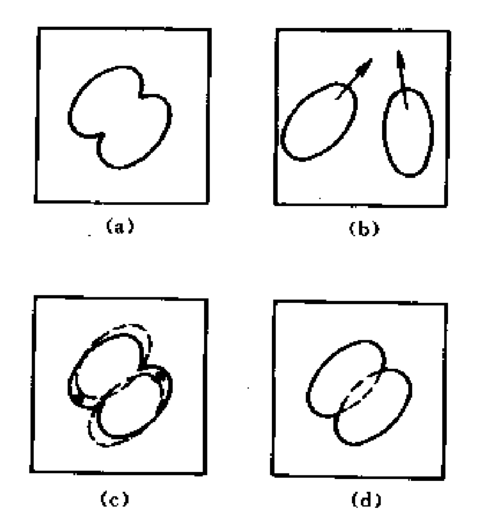


图 7－26　重叠物体的分离

(a)、(b)重叠前的物体和运动矢量

(c)虚线：利用(b)推断物体的位置

(d)重叠物体的分离

在运动图像的分析中，包含了从运动物体的检测、运动矢量的分析、运动物体的跟踪这些比较接近信号处理的方法，还有从物体形变或三维形状的识别以及行动的理解(运动因果关系的识别)这样一些需要进行复杂分析的方法。分析中最根本的是从多幅图像寻找对应于同一物体部分的匹配方法。可以使用从上面所讲的差分图像这样的简单方法，一直到求出图形结构之间的匹配这样较复杂的方法。

另外，因为运动图像的数据量很庞大，所以常采取利用从已经分析的图像上得到的信息，积极地用于以后图像的分析上，这是一种行之有效的方法。研究运动图像处理的历史还很短，有许多问题有待于今后的发展。

## 习　　题

7—1　设 $f_1$ 为对图像 $f$ 设置门限 $t$ 切割后的二值图像。$f_1$ 中两个灰度级为 $B$ 和 $S$。试证明不论 $f$ 中灰度级的概率密度如何，使积分 $\iint (f-f_1)^2 \mathrm{d}x\mathrm{d}y$ 为最小的 $t$ 值始终是 $t=(B+S)/2$

7－2　试用罗伯特梯度和拉普拉斯算子检测下图所示图像的边缘。

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 |
| 4 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 0 | 0 |
| 4 | 4 | 5 | 6 | 6 | 6 | 5 | 4 | 0 | 0 |
| 4 | 4 | 5 | 6 | 7 | 6 | 5 | 4 | 0 | 0 |
| 4 | 4 | 5 | 6 | 6 | 6 | 5 | 4 | 0 | 0 |
| 4 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 0 | 0 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 |

7－3　已知一个原始图像如图示，基于灰度级的值和灰度级的直方图，图中可以假定由两个区域组成，点(3,2)、(3,4)分别是假定区域的起始点。请用计算欧几里得距离的方法将图像进行分割，请计算出欧几里得距离为3和4两种情况下的分割结果。

|  | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 4 | 6 | 7 |
| 2 | 1 | 1 | 5 | 8 | 7 |
| 3 | 0 | ⊥ | 6 | ⊥ | 7 |
| 4 | 2 | 0 | 7 | 6 | 6 |
| 5 | 0 | 1 | 4 | 6 | 4 |

# 参 考 文 献


[7－1]　Duda Ro, Hart Pe, Use of Hoagh transformation to detect lines and earres in Picture Comm. ACM 15(1):11—15 1972

[7－2]　Tsuku Het al, Extracting elliptical figures from an edge field. Proc, of IEEE Computer vision and Pattern Recognition 1983 138—141

[7－3]　横矢，富田，田村"图像处理フルユ"リア"ムσ概观(5)—邻域分割"电子技术总合研究所汇款 Vol. 44, No, 718, PP. 96—112, 1980

[7－4]　田村秀行等著。计算机图像处理技术

[7－5]　荆仁杰等编著。计算机图像处理。浙江大学出版社，1990

# 第八章　图像的匹配

在机器识别事物的过程中，常需把不同传感器或同一传感器在不同时间，不同成像条件下对同一景物获取的两幅或多幅图像在空间上对准，或根据已知模式到另一幅图中寻找相应的模式，这就叫做匹配。在遥感图像处理中需把不同波段传感器对同一景物拍得多光谱图像按象点对应一一套准，然后根据象点的性质进行地物分类。如果利用在不同时间对同一地面拍摄得两幅照片，经套准后找出其中那些特征有了变化的象点，就可以用来分析图中哪些部分发生了变化，而利用放在一定间距处的两只传感器对同一物体摄得的两幅照片，找出对应点后可算出物体离开摄像机的距离，即深度信息，这叫体视学。其它如对序列图像匹配求光流场，以它描述三维动态景物，计算刚体的空间结构和运动参量等，凡此种种，匹配技术都起着至关重要的作用，自然地受到人们的很大重视，并已提出众多的匹配方法。

早期的匹配技术主要用于几何校正后的多波段遥感图像的套准，借助于求互相关函数的极值来实现。但在三维景物分析中，由于三维成象中有透视失真、运动遮挡，阴影混入和噪声干扰等不利因素，使得三维图像匹配至今仍是公认的技术难题，下面就一些代表性的方法作一些分析。

## §8.1　模板匹配法

### 一、模板匹配法

前面提到过的模板匹配是对图像进行边缘锐化及检测的方法，这里讨论图像的匹配，亦可以用模板匹配法，但是研究在一幅图中是否存在某种已知模板图像，例如在图 8－1(a)中，需要寻找一下有无三角形(图 8－1(b))的图像？若在被搜索图中有待寻的目标，且同模板有一样的尺寸和方向，它的基本原则就是通过相关函数的计算来找到它以及被搜索图的坐标位置。

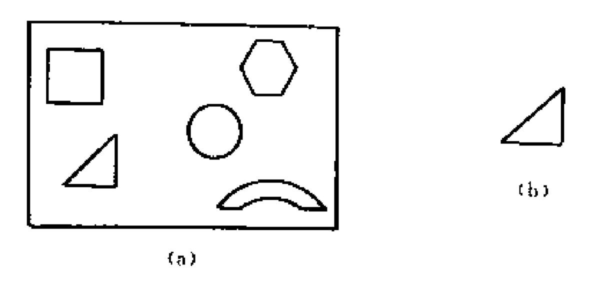


图 8－1　模板(b)与被搜索图(a)

设模板 $T$ 叠放在搜索图 $S$ 上平移，模板覆盖下的那块搜索图叫做子图 $S^{i,j}$，$i,j$ 为这块子图的左上角象点在 $S$ 图中的坐标，叫参考点，不难从图 8－2 中看出，$i$ 和 $j$ 的取值范围为

$$1 \leqslant i,j \leqslant N-M+1$$

现在可以比较 $T$ 和 $S^{i,j}$ 的内容，若两者一致，则 $T$ 和 $S^{i,j}$ 之差为零。所以可以用下列两种测度之一来衡量 $T$ 和 $S^{i,j}$ 的相似程度。

$$D(i,j)=\sum_{m}\sum_{n=1}^{M}[S^{i,j}(m,n)-T(m,n)]^2 \tag{8.1.1}$$

或

$$D(i,j)=\sum_m\sum_{n=1}^{M}|S^{i,j}(m,n)-T(m,n)| \tag{8.1.2}$$

若展开式 8.1.1，则有

$$D(i,j)=\sum_m\sum_n[S^{i,j}(m,n)]^2-2\sum_m\sum_n S^{i,j}(m,n)\cdot T(m,n)+\sum_m\sum_n[T(m,n)]^2 \tag{8.1.3}$$

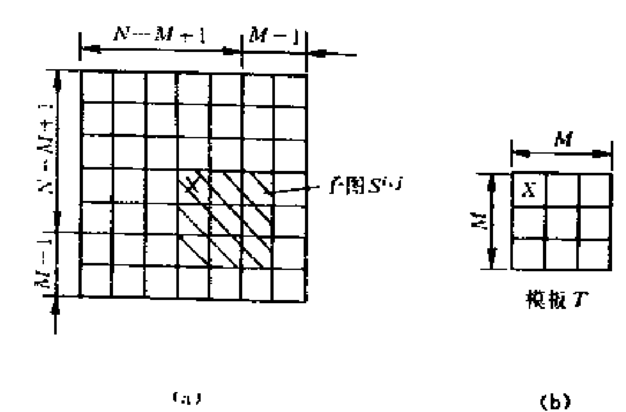



图 8-2　模板(b)及其搜索图(a)

式右边第三项表示模板的总能量，是一个常数，与$(i,j)$无关，第一项是模板覆盖下那块图像子图的能量，它随$(i,j)$位置而缓慢改变，第二项是子图像和模板的互相关，随$(i,j)$而改变，$T$ 和 $S^{i,j}$匹配时这项取值最大，因此可用下列相关函数作相似性测度：

$$R(i,j)=\frac{D\text{的第二项}}{D\text{的第一项}}=\frac{\sum_m\sum_n S^{i,j}(m,n)\cdot T(m,n)}{\sum_m\sum_n[S^{i,j}(m,n)]^2} \tag{8.1.4}$$

或归一化为

$$R(i,j)=\frac{\sum_m\sum_n S^{i,j}(m,n)\cdot T(m,n)}{(\sum_m\sum_n[S^{i,j}(m,n)]^2)^{1/2}(\sum_m\sum_n[T(m,n)]^2)^{1/2}} \tag{8.1.5}$$

根据施瓦兹不等式可以知道式(8.1.5)中的，$0\leqslant R(i,j)\leqslant 1$，并且仅在比值 $S^{i,j}(m,n)/T(m,n)$为常数时 $R^{i,j}$取极大值(等于 1)。式(8.1.5)可写成更简洁的内积形式，令 $S_1(i,j)$表示子图，$t$ 表示模板，则有

$$R(i,j)=\frac{t^TS_1(i,j)}{(t^Tt)^{1/2}[S_1^T(i,j)S_1(i,j)]^{1/2}} \tag{8.1.6}$$

当矢量 $t$ 和 $S_1$ 之间的夹角为零时，即 $S_1(i,j)=kt$ 时，这里 $k$ 为标量常数，有 $R(i,j)=1$，否则 $R(i,j)<1$。用相关法求匹配的计算量很大，因为模板要在$(N-M+1)^2$ 个参考位置上做相关计算，其中除一点以外都是在非匹配点上做无用功。因此，人们希望有一种快速计算方法，人们提出一类叫序贯相似性检测的算法，简称 SSDA，其要点是：

(1)定义绝对误差值：

$$\varepsilon(i,j,m_k,n_k)=|S^{ij}(m_k,n_k)-\hat{S}(i,j)-T(m_k,n_k)+\hat{T}|$$

式中

$$\hat{S}(i,j)=\frac{1}{M^2}\sum_{m,n=1}^{M}\sum S^{i,j}(m,n)$$

$$\hat{T}=\frac{1}{M^2}\sum_{m,n=1}^{M}\sum T(m,n)$$

(2)取一不变阈值 $T_k$。

(3)在子图 $S^{i,j}(m,n)$中随机选取象点，计算它同 $T$ 中对应点的误差值 $\varepsilon$，然后把这差值同其它点对的差值累加起来，当累加 $r$ 次误差超过 $T_k$，则停止累加，并记下次数 $r$，定义 SSDA 的检测曲面为

$$I(i,j)=\left\{r\left|\min_{1\leqslant r\leqslant m^2}\left[\sum_{k=1}^{r}\varepsilon(i,j,m_k,n_k)\geqslant T_k\right]\right.\right\} \tag{8.1.7}$$

(4)把$I(i,j)$值大的$(i,j)$点作为匹配点，因为这点上需要很多次累加才使总误差$\Sigma\epsilon$超过$T_k$，见图(8-3)，图中给出了在$A$、$B$、$C$三参考点上得到的误差累计增长曲线。$A$、$B$反映模板$T$不在匹配点上，这时$\Sigma\epsilon$增长很快，超出阈值，曲线$C$中$\Sigma\epsilon$增长很慢，很可能是一套准的候选点。

对SSDA算法还可以进一步改进计算效率，办法是：

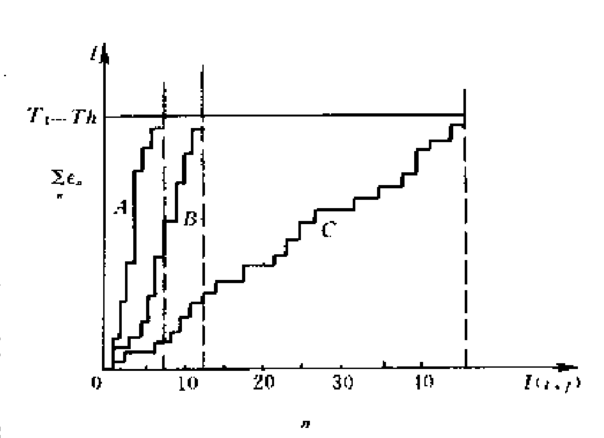



图8-3　$T_k$=常数时的累计误差增长曲线

(1)对于$(N-M+1)^2$个参考点的选用顺序可以不逐点推进，即模板不一定对每点都平移到，例如可用粗—细结合的均匀搜索，即先每隔$m$点搜一下匹配好坏，然后在有极大匹配值周围的局部范围内对各参考点位置求匹配，这策略能否不丢失真正匹配点，将取决于表面$I(i,j)$的平滑性和单峰性。

(2)在某参考点$(i,j)$处，对模板覆盖下的$M^2$个点对的计算顺序可用与$i$、$j$无关的随机方式计算误差，也可采用适应图像内容的方式，按模板中突出特征选取伪随机序列，决定计算误差的先后顺序，以便及早抛弃那些非匹配点。

(3)模板在$(i,j)$点得到的累计误差映射为上述曲面数值的方法，是否最佳还可以探索。

(4)不选用固定阈值$T_k$，而改用单调增长的阈值序列，使非匹配点用更少的计算就达到阈值而被丢弃，真匹配点则需更多次误差累计才达到阈值(见图8-4)。

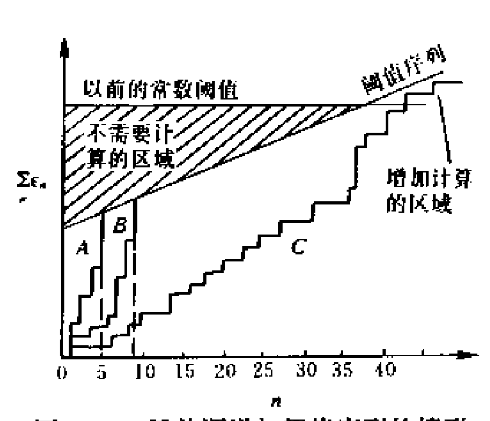



图8-4　用单调增加阈值序列的情形

由于除一点以外，绝大多数的情况下都是对非匹配点计算的，显然，越早丢弃非匹配点将越节省时间，Barnea等人给出了选用这种阈值序列为计算式子[8-1]。

SSDA方法比用FFT做相关法快50倍，是较受人重视的一种算法。对于二值图，SSDA还可简化，这是模板与对应子图中的成对象点的差值为

$$|S^{i,j}(m,n)-T(m,n)|=\bar{S}^{i,j}T+S^{i,j}\bar{T}$$
$$=S^{i,j}(m,n)\oplus T(m,n)$$

式中$\oplus$表示异或处理(模2加)，由此得到

$$D(i,j)=\sum_{m}^{M}\sum_{n}S^{i,j}(m,n)\oplus T(m,n) \tag{8.1.8}$$

这常被称为二进制的Hamming距离，$D$越小，则子图同模板越相似。

## 二、模板匹配法的改进

用模板进行匹配处理时，希望一要匹配点准确(相关峰尖锐，定位精度高)，二要计算量小。在用简单的相关性公式(8.1.5)计算时所得到的相关曲线往往较平坦，这因为第一，未考虑图像中各象点间存在空间相关性，第二，噪声可能掩盖相关峰值。为克服这缺点，可用'白化'滤波方法。

设图像中有噪声，写成矢量形式有

$$f_n=f+n \tag{8.1.9}$$

式中，$f$ 为确定性图像，$n$ 为噪声。今用匹配滤波器 $m$ 对 $f$ 过滤，其输出为标量

$$f_0=m^T f_n \tag{8.1.10}$$

无噪声时，信号功率为 $S=(m^T f)^2$。而噪声功率为

$$N=E\{(m^T n)(m^T n)^T\}=m^T K_n m$$

式中 $K_n$ 为噪声的协方差矩阵，所以信噪比为

$$\frac{S}{N}=\frac{(m^T f)^2}{m^T K_n m} \tag{8.1.11}$$

今求最佳 $m$，它使 $S/N$ 最大。这可把上式对 $m$ 取导，使等于零，得到

$$m=\left[\frac{m^T K_n m}{m^T f}\right]K_n^{-1} f \tag{8.1.12}$$

式中方括号内是标量值，可归一化。这样，匹配滤波器的输出为 $f_0=f^T(K_n^{-1})^T f_n$。一般说噪声协方差矩阵可分解为

$$K_n=KK^T \tag{8.1.13}$$

式中 $K=E\Lambda_n^{1/2}$，$E$ 是本征矢量矩阵，$\Lambda_n$ 是本征值矩阵，代回有

$$f_0=[K^{-1}f][K^{-1}f_n]$$

称 $K^{-1}f_n$ 是把 $f_n$ 作“白化处理”。

将上述去相关的白化思想用到模板匹配中，对于子图 $S^{i,j}$ 和模板 $T$ 分别用矢量 $s(i,j)$ 和 $t$ 表示，这时有

$$g_1=H_1^{-1}t$$

$$g_2(i,j)=H_2^{-1}S(i,j)$$

式中 $\qquad H_k=E_k\Lambda_k^{1/2} \qquad K_k=H_k H_k^T \qquad K=1.2$

于是得到（统计）相关函数为

$$\begin{aligned}R_S(i,j)&=\frac{g_1^T g_2(i,j)}{[g_1^T g_1]^{1/2}[g_2^T(i,j)g_2(i,j)]^{1/2}}\\&=\frac{[(K_3^T)^{-1}t)]^T S_1(i,j)}{[(K_1^T)^{-1}t]^T t[(K_2^T)^{-1}S_1(i,j)]^T S_1(i,j)}\end{aligned} \tag{8.1.14}$$

设模板尺寸为 16×16，协方差矩阵 $K_1$、$K_2$ 都是 256×256 阶的，求其本征矢量和本征值的计算量都很大，实际应用中有困难。这时，可改用另一种统计相关函数，它定义为。

$$Rs(i,j)=\frac{\sum_m\sum_n G_1^{i,j}(m,n)G_2(m,n)}{\{\sum_m\sum_n[G_1^{i,j}(m,n)]^2\}^{1/2}\{\sum_m\sum_n[G_2(m,n)]^2\}^{1/2}} \tag{8.1.15}$$

式中

$$\left.\begin{aligned}G_1^{i,j}(mn)&=S^{i,j}(m,n)*D_1(m,n)\\G_2(m,n)&=T(m,n)*D_2(m,n)\end{aligned}\right\} \tag{8.1.16}$$

如用马尔可夫过程模拟图像，并不计噪声时，Pratt 认为，式(8.1.16)中滤波器冲激响应 $D_i(m,n)$ 可用下式表示

$$D_i=\begin{bmatrix}\rho_c\rho_R & -\rho_c(1+\rho_R^2) & \rho_c\rho_R\\-\rho_R(1+\rho_c^2) & (1+\rho_c^2)(1+\rho_R^2) & -\rho_R(1+\rho_c^2)\\\rho_c\rho_R & -\rho_c(1+\rho_R^2) & \rho_c\rho_R\end{bmatrix} \tag{8.1.17}$$

式中 $\rho_R$ 和 $\rho_c$ 分别代表相邻行和相邻列的象点的马尔可夫相关系数。若 $\rho_R=\rho_c=\rho$，则可化为

$$D_i=\begin{bmatrix} \rho^2 & -\rho(1+\rho^2) & \rho^2 \\ -\rho(1+\rho^2) & (1+\rho^2)^2 & -\rho(1+\rho^2) \\ \rho^2 & -\rho(1+\rho^2) & \rho^2 \end{bmatrix} \tag{8.1.18}$$

若 $\rho=1$，则有

$$D_i=\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

这是拉普拉斯微分算子，因此当图像中象点高度相关时，计算二图片中景物的相关性实际是集中在计算它们轮廓信息的相关性。图 8-5 是计算机的模拟结果，被搜索图取为 32×32，模板 16×16，水平方向和垂直方向分别偏移了三个和四个象点，用统计相关函数式(8.1.15)计算。$\rho=0$ 相当于普通相关(即用式(8.1.5)计算)，$\rho\geqslant0.9$ 时，$R_i(i,j)$的峰在套准点附近很尖锐。

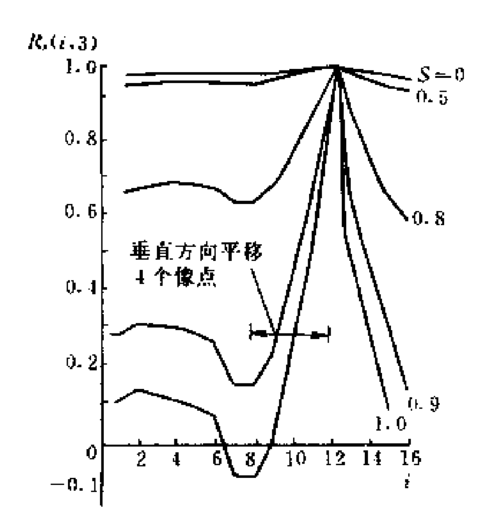



图 8-5 图像匹配的统计相关函数曲线

Wong 提出一种简单而有效的相似性相关法[8-2]，他引用了成对函数的概念，设模板 $T$ 和子图 $S^{i,j}$都由 $M\times M$ 点组成，则有 $M^2$ 个点对，点对中任一象点的灰度值 $\alpha$ 可在 0～($2^b$—1)间取值，这里 $b$ 是象点量化的比特数。定义成对函数 $N_{\alpha\beta}(i,j)$为子图中象点灰度值 $\alpha$，对应位置上模板中象点灰度值为 $\beta$ 的点对数目。例如，图 8-6 中 $N_{\alpha\beta}(i,j)=2$，显然，若子图同模板在灰度上完全匹配，则有

$$\sum_{\alpha=0}^{2^b-1}\sum_{\beta=0}^{2^b-1}N_{\alpha\beta}(i,j)=\begin{cases}M^2 & \alpha=\beta \\ 0 & \alpha\neq\beta\end{cases} \tag{8.1.19}$$

因此它定义成对函数的相关测度为

$$R_{PF}(i,j)=\prod_{\alpha=0}^{2^b-1}\left[N_{\alpha\alpha}(i,j)\Big/\sum_{\beta=0}^{2^b-1}N_{\alpha\beta}(i,j)\right] \tag{8.1.20}$$

方括号中分子是匹配点对数，分母是灰度值取可能的组合的点对数。例如 $b=2$，则灰度有 4 级(0,1,2,3)，可能的灰度组合点对数有 16 种，于是

$$R_{PF}(i,j)=\left(\frac{N_{00}}{N_{00}+N_{01}+N_{02}+N_{03}}\right)\left(\frac{N_{11}}{N_{10}+N_{11}+N_{12}+N_{13}}\right)\times \left(\frac{N_{22}}{N_{20}+N_{21}+N_{22}+N_{23}}\right)\left(\frac{N_{33}}{N_{30}+N_{31}+N_{32}+N_{33}}\right)$$

式(8.1.20)右边有 $2^b$ 个括号相乘，括号中分式的分母是 $2^b$ 项之和。对于二值图($b=1$)，$R_{PF}(i,j)$表达式极其简单，为

$$R_{pF}(i,j)=\frac{1}{1+(N_{01}/N_{00})}\times\frac{1}{1+(N_{10}/N_{11})}$$

只需用 4 个计数器累计 $N_{\alpha\beta}$值，就可算出 $R_{PF}(i,j)$，本相关测度的峰值为 1，背景值(偏离匹配点)较低，峰较尖锐，见图 8-7，图中所用搜索区大小为 32×32，模板为 8×8。

最后强调一下，为减少成对函数相关测度 $R_{pF}(i,j)$的计算量，图像中象点量化的灰度极不

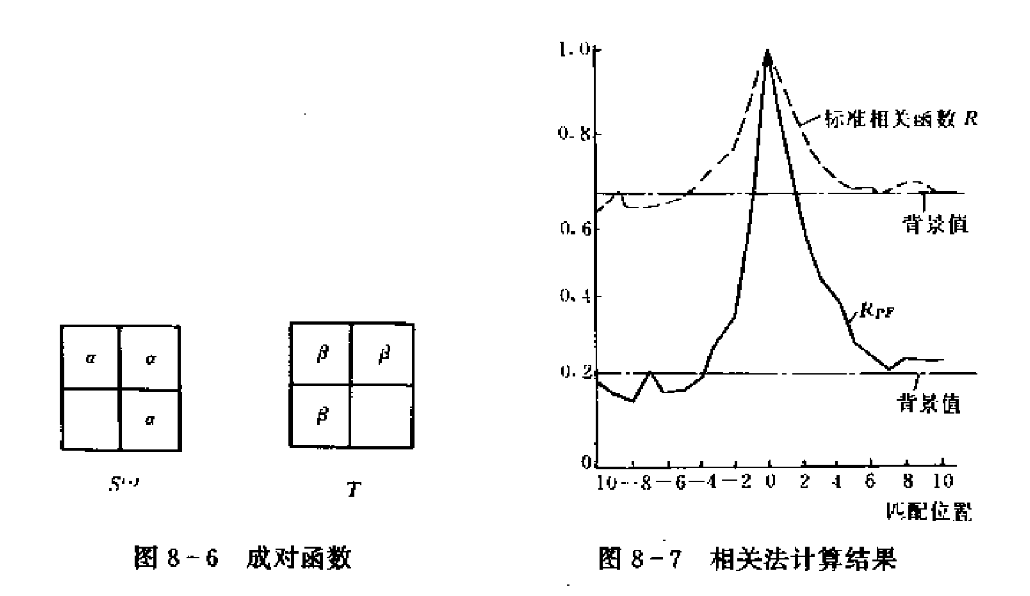



图 8-6　成对函数　　　　图 8-7　相关法计算结果

宜过多。

## §8.2　其它快速计算法

下面我们将以地形和地图匹配的技术为例，讨论加快匹配的其它算法，文中提到的实时图是指飞行器实时测量地面所得到的数据图。而飞行器预存在机内的已知地面的数据图叫做基准图，两图的匹配计算可以对地形进行分析，并可以指示出飞行器当前的位置。

由前面讨论可知，任何一种匹配算法的总的计算量由所采用的相关算法的计算量与搜索位置数之积，即

总计算量=(相关算法的计算量)×(搜索位置数)

来决定的。因此，为了减少总的计算量，除上面介绍的方法以外我们再介绍几种方法

### 一、幅度排序相关算法

这种算法由两个步骤组成，第一步把实时图中的各个灰度值按幅度大小排成列的形式，然后再对它进行二进制(或三进制)编码，最后，根据二进位排序的诸列，把实时图变换成二进制阵列的一个有序的集合$\{C_n, n=1,2\cdots N\}$。这一过程称之为幅度排序的预处理。第二步，序贯地将这些二进制阵列与基准图进行由粗到细的相关，直到确定出匹配点为止。这里，为了说明这种算法的原理，举一个简单的 3×3 实时图的例子。

第一步：预处理。

首先把 3×3 实时图中各个灰度值按大小次序排成一列，并算出各个灰度值在原图中的位置$(j,k)$，如图 8-8 所示。

然后把排序后的灰度幅度值分成数目相等的两组，且幅度大的一组赋值为 1，而幅度小的一组赋值为 0，若幅度数为奇数，则中间的那个幅度就规定为×，如图 8-8(b)所示。进一步，把每一组分成两半，并同样地赋予 1 值和 0 值。这个过程一直进行到各组划分为一个单元为止，并由此形成二进制排序。

|  | k |  |  |
|---|---|---|---|
| j | 1 | 2 | 3 |
| 1 | 11 | 17 | 1 |
| 2 | 15 | 6 | 5 |
| 3 | 14 | 4 | 10 |

(a)3×3 实时图

| 位置 (j k) | 幅度 | 二进制排序 ① | ② | ③ |
|---|---|---|---|---|
| 1,2 | 17 | 1 | 1 | 1 |
| 2,3 | 15 | 1 | 1 | 0 |
| 3,1 | 14 | 1 | 0 | 1 |
| 1,1 | 11 | 1 | 0 | 0 |
| 3,3 | 10 | × | × | × |
| 2,1 | 6 | 0 | 1 | 1 |
| 2,2 | 5 | 0 | 1 | 0 |
| 3,2 | 4 | 0 | 0 | 1 |
| 1,3 | 1 | 0 | 0 | 0 |

(b)预处理步骤(排序)

图 8-8　3×3 实时图的预处理

于是，根据二进制排序的次序①、②、③和各个二进制值及其位置，便可构成 $C_1,C_2,C_3$ 等二进制阵列，如图 8-9 所示。同理，对于一般情况可得 $C_n, n=1,2\cdots N$，此处，$N$ 为二进制排序的分层数。

| 1 | 1 | 0 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0 | × |

$C_1$

| 0 | 1 | 0 |
|---|---|---|
| 1 | 1 | 1 |
| 0 | 0 | × |

$C_2$

| 0 | 1 | 0 |
|---|---|---|
| 1 | 0 | 0 |
| 1 | 1 | × |

$C_3$

图 8-9　本例的二进制阵列

第二步：由粗到细的相关过程。

首先，用 $C_1$ 阵列与基准图阵列作如下相关运算，得

$$\varphi_1(u,v)=\sum_{\substack{j,k\\C_1(j,k)=1}} X_{j+u,k+v}-\sum_{\substack{j,k\\C_1(j,k)=0}} X_{j+u,k+v} \tag{8.2.1}$$

这式意味着，当 $C_1$ 阵列放在基准图的某一搜索位置 $(u,v)$ 上时，与 $C_1$ 中的 1 值所对应的基准图的象元值之和减去与 $C_1$ 中的 0 值所对应的基准图的象元值之和（而与 $C_1$ 中×号所对应的基准图像元，则被忽略之）。所以 $\varphi_1(u,v)$ 实际上是一比特量化实时图与基准图的积相关函数，它反映了实时图中最粗糙的图像结构的信息（即一种高低表示）与基准图的相关。称 $\varphi_1(u,v)$ 为基本的相关面。

在基准图全区域的搜索过程中，若设定一个门限值 $T_1$，并舍弃那些 $\varphi_1(u,v)<T$ 的试验点，则就可以大大减少下一轮搜索时的试验位置数。

然而，在 $\varphi_1(u,v)>T_1$ 的试验位置上，再进行细的相关运算，这可以用下式来计算。

$$\varphi_2(u,v)=\varphi_1(u,v)+\frac{1}{2}\Big\{\sum_{\substack{j,k\\C_2(i,j)=1}} X_{j+u,k+v}-\sum_{\substack{j,k\\C_2(j,k)=0}} X_{j+u,k+v}\Big\} \tag{8.2.2}$$

同理，为了减少有争议的匹配点的数目，设门限值 $T_2$，并在 $\varphi_2(u,v)>T_2$ 的试验位置上，以 $C_2$ 为基础进行更细的相关运算

$$\varphi_3(u,v)=\varphi_2(u,v)+\frac{1}{2^2}\Big\{\sum_{\substack{j,k\\C_3(i,k)=1}} X_{j+u,k+v}-\sum_{\substack{j,k\\C_3(j,k)=0}} X_{j+u,k+v}\Big\}$$

再设门限 $T_3$ 等等，依此类推，可得第 $n$ 个相关面为

$$\varphi_n(u,v)=\varphi_{n-1}(u,v)+\frac{1}{2^{n-1}}\Big\{\sum_{\substack{j,k\\C_n(j,k)=1}}X_{j+u,k+v}-\sum_{\substack{j,k\\C_n(j,k)=0}}X_{j+u,k+v}\Big\} \tag{8.2.3}$$

当设门限值为 $T_n$ 时，若 $\varphi_n(u^*,v^*)>T_n$ 的位置只有一个，就宣布该位置 $(u^*,v^*)$ 为匹配位置。

显然，各个门限值有如下的关系

$$T_n>T_{n-1}\cdots>T_3>T_2>T_1$$

因此，逐次细化相关的试验位置将越来越少，直到找出匹配位置时为止。从而减少了总的计算量，亦即提高相关的处理速度。

这种二进制位的幅度排序算法(即 BARC 算法)，所需要的加法总计算量为

$$q=k(M_1-N_1+1)(M_2-N_2+1)N_1N_2 \tag{8.2.4}$$

其中 $k$ 是一个在 1 和 2 之间的常数，而 $M_1$、$M_2$ 和 $N_1$、$N_2$ 分别为基准图和实时图的尺寸。

## 二、FFT 的相关算法

由傅里叶分析中的相关定理可知，两个函数在定义域中的卷积等于它们在频域中的乘积，而相关则是卷积的一种特定形式。因之，存在着另一种计算相关函数的方法。但是这样做在时间上并没有多少可取之处，但由于快速傅里叶变换技术比直接法计算速度提高了一个数量级，为此用 FFT 进行频域相关计算亦是一种可行的方法。

首先把基准图和实时图进行二维离散傅里叶变换(DFT)，对于基准图，有

$$X(u,v)=\frac{1}{M^2}\sum_{j=0}^{M-1}\sum_{k=0}^{M-1}x(j,k)\omega_M^{-uj}\omega_M^{-vk} \tag{8.2.5}$$

其中 $u$ 和 $v$ 分别表示 $j$ 和 $k$ 方向上的频率变量，且

$$\omega_M\triangleq\exp(\mathrm{j}\,\frac{2\pi}{M})$$

并假定基准图的尺寸是 $M\times M$ 维的。实时图的离散傅里叶变换 $Y(u,v)$ 是用同样的方式计算得到的。然后，由相关定理可以写出相关的离散傅里叶变换 $\phi(u,v)$ 为

$$\phi(u,v)=X(u,v)Y^*(u,v)$$

为此，对 $\phi(u,v)$ 求傅里叶反变换，就可以得出空间域中的相关函数 $\phi(j,k)$ 为

$$\phi(j,k)=\sum_{u=0}^{M-1}\sum_{v=0}^{M-1}[X(u,v)*Y^*(u,v)]\omega_M^{uj}\omega_M^{vk} \tag{8.2.6}$$

其中 * 为共轭号

由此可见，相关函数可以通过 DFT 的方法计算出来。而计算 DFT 最有效的方法就是采用 FFT 算法，在一般的教科书上都可以找到。所以这里略去对它的讨论。

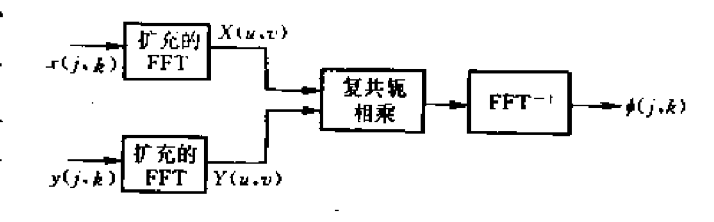



图 8-10 FFT 相关算法

最后根据以上的关系式，我们可以画出 FFT 的相关算法流程图，如图 8-10 所示。显然，这种相关算法可以容易地推广到 FFT 的归一化相关算法，感兴趣的读者可参阅文献[8-3]。

如果被测试图像的象元素和试验位置数越大的话，那么，相对地说，这种算法在时间上的节省就更大。但是，要注意，由于傅里叶变换是个周期性函数，因此匹配点会周期地出现，所以在运算时，必须采取其它适当的措施才行。

### 三、分层搜索的序贯判决算法

这种分层搜索算法是直接基于人们先粗后细寻找事物的贯例而引成的，例如，在世界地图上找北京的位置时，可以先找出中国这个广阔的地域，称其为粗相关，然后在这个地域内，再仔细确定北京的位置，这叫做细相关。很明显，利用这种找法，就可以很快地找到北京的位置。所以由这种思想形成的分层搜索算法具有相当高的处理速度。如图 BARC 算法一样，它也是由二个步骤组成的。

第一步：预先处理。

首先，对被匹配的图像进行分层预处理，方法是将图 2×2 维的邻区逐个网络地进行平均处理，从而得到一个分辨力较低和维数较小的图像。然后，将此图像再用同样的方法处理后，得到一个分辨力更低和维数更小的图像。依次进行下去，如果一共进行了 $K$ 次分层处理的话，那么我们就可以得到 $K$ 个处理后的图像。于是，加上原图像后，便可构成一组分辨力由高到低，而维数由大到小的图像序列。这种技巧称之为分层预处理。

如果将上述分层技术应用于基准图和实时图，则就可以获得两个这样的图像序列，一个是基准图的，另一个是实时图的，它们分别表示为

$$X_k:(\frac{M}{2^k}\times\frac{M}{2^k})$$

$$Y_k:(\frac{N}{2^k}\times\frac{N}{2^k}) \tag{8.2.7}$$

其中 $k=0,1,2,\cdots,L$，且已假设基准图是 $M\times M$ 维的，而实时图是 $N\times N$ 维的。

因为原图的分辨力最高，且维数最大，所以 $k=0$ 的图像 $X_0$ 和 $Y_0$ 具有最高的分辨力，而 $k=L$ 的图像 $X_L$ 和 $Y_L$，则具有最低的分辨力。

如前所述，若采用先粗后细的相关方法，则可以很快地找到匹配点。所以，第一次相关搜索是从分辨力最低和维数最小的一对图像 $X_K$、$Y_K(K=L)$开始的。这时，由于 $Y_K$ 的象元数比较少，加上损失了一部分高频信息，所以，在粗相关的过程中，正确截获概率 $P_C^K$ 将是不大的，所以，为了提高 $P_C^K$ 应设法改善 $X_K$ 和 $Y_K$ 的信噪比，例如，将较高分辨力的 $Y_{K-1}$(或 $X_{K-1}$)通过低通滤波器以后，再以二倍于它的空间采样间隔进行采样，而得到的 $Y_K$(或 $X_K$)比用直接分层法得到的具有更高的信噪比。因此，相对地说，提高了 $P_C^K$，显然，其它各层，亦作同样的处理。这种技术称之为分层搜索预处理。

第二步：先粗后细的相关过程。

如前所述，第一次相关是从 $Y_K$ 和 $X_K$ 开始的，为了找到可能的粗匹配位置，应将 $Y_K$ 在 $X_K$ 的所有搜索位置上进行相关，并确定出粗匹配位置($(u^k,v^k)$)。因为这时 $Y_K$ 和 $X_K$ 的维数最小，所以搜索过程是很快的，但是这时 $P_C^K$ 值较小，故可能产生若干个可能的粗匹配位置。第二次相关是在较高分辨力的图像 $Y_{k-1}$和 $X_{k-1}$之间进行的。这时，因为已经知道了可能的粗匹配位置，所以 $Y_{k-1}$只需要在 $X_{k-1}$的一个或若干个粗匹配位置附近进行相关搜索就可以了，从而找出一个或少数几个可能性更大的匹配位置$(u^{k-1},v^{k-1})$。在上述相关过程中，为了不致于丢失匹配点，应在粗匹配位置$(u^k,u^k)$附近增加几个补充的试验位置。显然，第三次相关与第二次相关是类似的，如此进行下去，一直到最高分辨力的实时图 $Y_0$ 在基准图 $X_0$ 上找到匹配位置时为止。由此可见，整个搜索过程是从最低分辨力到最高分辨力一层一层地进行下去的，为了进一步提高处理速度，相关运算常常采用 SSDA 算法。这种技术称之为分层搜索的序贯判决算法。

1. 搜索位置数

由上述讨论可知，除了在分辨力最低和尺寸最小的图上作全区域搜索之外，其它各层的搜索都是在少数几个可能匹配的位置上进行的。因此，由式(8.2.7)可知，当最低分辨力的两图进行相关时，总的搜索位置数为

$$\left(\frac{M}{2^L}-\frac{N}{2^L}+1\right)^2 \approx (M-N)^2/2^{2L} \tag{8.2.8}$$

而当最高分辨力($L=0$)的两图相关时，搜索位置数则为

$$(M-N+1)^2 \approx (M-N)^2$$

因此，如果不考虑其它各层的搜索位置数(很少)的话，那么分层搜索算法的搜索位置为一般算法的 $1/2^{2L}$，从而大大地提高了处理速度。

实验表明，当用分层搜索算法时，搜索位置数只有 $K\lg(M-N-1)^2$ 个，其中 $K=1\sim2$。

2. 分层搜索序贯判决算法的门限序列

这里，假定第 $K$ 次搜索级(即第 $K$ 分层)的低分辨力的图像是由第($K-1$)搜索级较高分辨力的图像通过低通滤波器以后，再以二倍于它的间隔采样后得到的。因此，如果低通滤波器的频率特性比较理想的话，那末由采样定理可知，这相当于把低通滤波器的噪声通带减小到1/2，换言之，从第 $K$ 级搜索到第 $K-1$ 级时，图像噪声增加了 $\sqrt{2}$ 倍。

如果采用 SSDA 算法，若采用均值一偏差门限序列，则其门限序列公式为

$$T_n^k=(\sqrt{2})^{L-k}(n+g_k\sqrt{n})r_L \tag{8.2.9}$$

其中 $T_n^k$ 表示第 $K$ 搜索级的判决门取序列，$k=0,1,2,\cdots,L$。$r_L$ 是在最低分辨力的 $L$ 级搜索级匹配位置上求得的噪声绝对值的均值，而 $g_k$ 是由搜索级 $K$ 的匹配概率 $P_D^k$ 决定的。

应该指出，对于同样的分层搜索技术，若采用不同的滤波预处理和不同的相关算法，则就可以形成不同的分层搜索的匹配算法，例如采用对函数的积相关算法，则可以形成分层搜索的对函数匹配算法[8-4]等。

## §8.3 受几何失真等影响小的匹配算法

如前所述，几何失真是影响图像匹配性能的一个重要的因素。因此，如何减少几何失真对匹配性能的影响，这是一个在实际中必须要加以考虑的问题。因此，在预处理方面，可以采用几何校正方法。或者，采用受几何失真等影响小的匹配算法，例如不变矩匹配算法，相位相关算法，多子区相关的算法及特征匹配算法等等，这些就是本节中所要讨论的问题。

### 一、不变矩匹配算法

在前面讨论图像描绘时曾经说过，可以求出一个图像的 7 个不变矩。这些不变矩当比例因子小于 2 和旋转不超过 45°的条件下，对于平移、旋转和比例因子的变化都是不变的，所以它反映了图像的固有的特性。因此，两个图像之间的相似度可以用它们 7 个不变矩之间的相似性来描述。这样的算法就称之为不变矩匹配算法(简称 IM 算法)，显然，在上述指出的条件下，它是不受几何失真影响的。

现在，如果令实时图的不变矩为 $M_i$，$i=1,2,\cdots,7$，并令试验位置$(u,v)$上的基准子图的不变矩为 $N_i(u,v)$，$i=1,2,\cdots,7$，则两图之间的相似度可以用任一种相关算法来度量。例如，当

采用归一化相关算法时有

$$R(u,v) = \frac{\sum_{i=1}^{7} M_i N_i(u,v)}{\left[\sum_{i=1}^{7} M_i^2 \sum_{i=1}^{7} N_i^2(u,v)\right]^{1/2}} \tag{8.3.1}$$

其中 $R(u,v)$ 是试验位置 $(u,v)$ 上的不变矩的相关值。

如前一样，取 $\max(R(u^*,v^*))$ 所对应的试验位置 $(u^*,v^*)$ 作为匹配点。显然，这种算法在进行相关之前，需要计算 7 个不变矩(预处理的一种形式)，所以，若采用常规的搜索方法，则要求较大的计算量。因此，为了提高它的处理速度，常常采用分层搜索技术，一般说，最低的搜索级取为 3 就可以了，因为太低的搜索级会影响到不变矩的计算精度。

## 二、相位相关算法

因为几何失真对图像的高频分量影响大，而对低频分量影响小，所以采用低通滤波器的，以傅里叶频谱为基础的相位相关算法，可以大大减小几何失真对匹配性能的影响。这种算法简叙如下：

若基准子图 $X_{u,v}$ 和实时图 $y$ 的离散傅里叶变换分别为 $X_{u,v}$ 和 $Y$，则它们的互功谱为

$$S_{u,v} = X_{u,v} Y^*$$

其中 * 为共轭符号。归一化后，便得到对应于这个互功谱的相位谱，为

$$e^{jQ_{uv}} \triangleq \frac{X_{u,v}Y^*}{|X_{u,v}Y^*|} = e^{j(Q_x - Q_y)} \tag{8.3.2}$$

其中 $Q_x$ 和 $Q_y$ 分别是基准子图 $X_{u,v}$ 和实时图 $Y$ 的傅里叶变换的相位。

由此及傅里叶变换的理论可知，这个相位谱包含了两图之间位置平移的所有信息，而且它是一个频谱幅度在全频域内为 1 的功率谱，所以，它是一个"白化"功率谱。

由式(8.3.2)可以看出，相位相关函数是一个位于两图位置偏移 $(u,v)$ 处的 $\delta$ 脉冲函数，在理想的情况下，当两图完全相似时，其值为 1，反之则为 0，因此，相位相关函数可以来度量两图之间的相似程度。由此形成的匹配算法，我们称之为相位相关算法。由于相位相关面 $\phi(u,v)$ 的形状类似于 $\delta$ 函数的形状，因此，这种算法具有较高的匹配精度。此外，还由于相位相关函数对于灰度值及其刻度的变化是不变的，所以，相位相关算法是不受这些误差因素影响的。

由于几何失真主要影响图像的高频成分，而大多数图像的频谱能量又集中在低频区域，因此，基于这一特点，若在频域中用一个低通滤波器函数 $H(f)$ 加权相位谱的话，则可以起到抑制几何失真和测量噪声的双重作用，这时，相位相关函数变为

$$\phi(u,v) = F^{-1}\{He^{jQ_{u,v}}\} \tag{8.3.3}$$

由这个相位相关函数构成的匹配算法，我们称之为带低通滤波器的相位相关算法。这时，由于只利用了图像的低频信息，所以它的匹配精度将有所下降。

顺便指出，如果把式(8.3.3)中的 $H$ 视为一个加权函数，并取如下的形式

$$H \triangleq |X_{u,v}Y^*|^{\alpha}$$

则有

$$\Phi_{(u,v)} = F^{-1}\{X_{u,v}Y^* |X_{u,v}Y^*|^{\alpha-1}\} \tag{8.3.4}$$

由此可见，当 $\alpha=0$ 时，$\phi(u,v)$ 代表相位相关函数，而 $\alpha=1$ 时，$\phi(u,v)$ 就是普通的相关函数，因此，通过 0 和 1 之间，改变 $\alpha$ 值，就可以产生两种匹配算法。

# 第九章　模式识别技术

前面我们已经介绍了图像的变换、增强、复原、匹配等技术，它们都是对输入图像的某种有效的改善，其输出仍然是一幅完整的图像。

随着数字图像处理技术的发展和实际应用的需求，出现了另一类问题，就是不要求其结果输出是一幅完整图像的本身，而是将经过上述处理后的图像，再经过分割和描述提取有效的特征，进而加以判决分类。例如要从遥感图像中分割出各种农作物、森林资源、矿产资源等等，并进一步判断其产量或蕴藏量，由气象云图结合其它气象观察数据进行自动天气预报，用人工地震波形图寻找有油的岩层结构，根据医学 X 光图像断层分析各种病变，邮政系统中的信函自动分拣等等。因此，可以认为把图像进行区别分类就是图像的模式识别。模式识别方法和应用十分广泛，也相当复杂，正在发展之中。模式识别的研究对象基本上可概括为二大类：一类是有直觉形象的如图像、相片、图案、文字等等，一类是没有直觉形象而只有数据或信息波形如语声、心电脉冲、地震波等等。但是，对模式识别来说，无论是数据、信号还是平面图形或立体景物都是除掉他们的物理内容而找出它们的共性，把具有同一共性的归为一类，而具有另一种共性者归为另一类。模式识别研究的目的是研制能够自动处理某些信息的机器系统，以便代替人完成分类和辨识的任务。

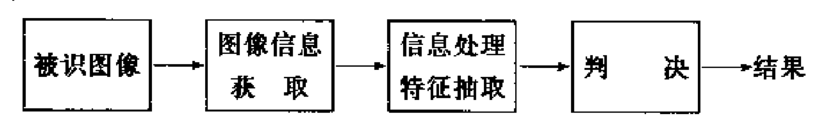



图 9-1　图像识别的简单框图

一个图像识别系统可分为三个主要部分，其框图如图 9-1 所示。第一部分是图像信息的获取，它相当于对被研究对象的调查和了解，从中得到数据和材料，对图像识别来说就是把图片、底片、文字图形等用光电扫描设备变换为电信号以备后续处理。第二部分是信息的加工与处理。它的作用在于把调查了解到的数据材料进行加工、整理、分析、归纳以去伪存真，去粗取精，抽出能反映事物本质的特征。当然，抽取什么特征，保留多少特征与采用何种判决有很大关系。第三部分是判决或分类。这相当于人们从感性认识升到理性认识而做出结论的过程。第三部分与特征抽取的方式密切相关。它的复杂程度也依赖于特征的抽取方式。例如，类似度、相关性、最小距离等等。

模式识别已初步形成三大类：统计模式识别、结构模式识别、模糊集识别等三大类，下面将介绍统计模式识别法。

## §9.1　统计模式识别法

统计模式识别的过程如图 9-2 所示，这是计算机识别的基本过程。数字化的任务是把图像信号变成计算机能够接受的数字信号。预处理的目的是去除干扰、噪声及差异、将原始信号变成适合于计算机进行特征抽取的形式，然后对经过预处理的信号进行特征抽取。最后进行判决分类，得到识别的结果。为了进行分类，必须有图像样本。对样本图像进行特征选择及学习

是识别处理中所必要的分析工作。

## 一、决策理论方法

正如图 9-2 的框图所示，统计模式识别方法最终归结为分类问题。假如已抽取出 $N$ 个特征，而图像可分为 $m$ 类。那么就可以对 $N$ 进行分类，从而决定未知图像属于 $m$ 类中的哪一类。一般把识别模式看成是 $N$ 维空间中的向量 $X$，即

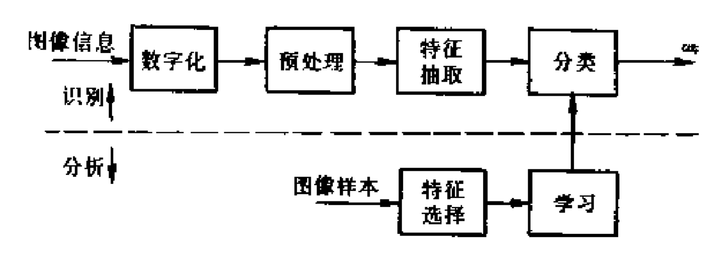



图 9-2 统计模式识别框图

$$X = [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_N]^T$$

模式类别为 $\omega_1$、$\omega_2$、$\omega_m$，识别就是要判断 $X$ 是否属于 $\omega_i$ 以及 $x_i$ 属于 $\omega_m$ 中的哪一类。在这个过程中主要解决两个问题：一是如何抽取特征，要求特征数 $N$ 尽可能小而且对分类判断有效，二是假设已有了代表模式的向量，如何决定它属于哪一类，这就需要判别函数。例如，模式有 $\omega_1$、$\omega_2$、$\cdots$、$\omega_m$ 共 $m$ 个类别，则应有 $D_1(X)$、$D_2(X)$、$D_3(X)$、$\cdots$、$D_m(X)$共 $m$ 个判别函数。如果 $X$ 属于第 $i$ 类，则有

$$D_i(X) > D_j(X) \qquad (j = 1,2,3,\cdots,m; j \neq i)$$

在两类的分界线上，则有 $\quad D_i(X)=D_j(X)$

这时 $X$ 既属于第 $i$ 类，也属于第 $j$ 类，因此这种判别失效。为了进行识别就必须重新考虑其它特征，再进行识别。问题的关键是找到合适的判别函数。

1. 线性判别函数

线性判别函数是应用较广的一种判别函数，所谓线性判别函数是指判别函数是图像所有特征量的线性组合，即

$$D_i(X) = \sum_{k=1}^{N} \omega_{ik} X_k + \omega_{io} \tag{9.1.1}$$

式中 $D_i(X)$代表第 $i$ 个判别函数，$\omega_{ik}$是系数或权，$\omega_{io}$为常数或称为阈值。在两类之间的判决界处有

$$D_i(X) - D_j(X) = 0 \tag{9.1.2}$$

该方程在二维空间是直线，在三维空间是平面，在 $N$ 维空间则是超平面。$D_i(X)-D_j(X)$ 可以写成以下的形式

$$D_i(X) - D_j(X) = \sum_{k=1}^{N} (\omega_{ik} - \omega_{jk}) X_k + (\omega_{io} - \omega_{jo})$$

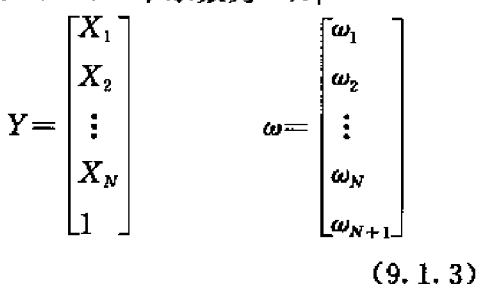
其判决过程可如下进行：如果 $D_i(X)>D_j(X)$，或 $D_i(X)-D_j(X)>0$，则 $X\sim\omega_i$，如果 $D_i(X)<D_j(X)$或 $D_i(X)-D_j(X)<0$，则 $X\sim\omega_j$。

用线性判别函数进行分类是线性分类器。任何 $m$ 类问题都可以分 解为$(m-1)$个 2 类识别问题。方法是先把模式空间分为 1 类和其他类，如此进行下去即可。因此，两类线性分类器是最简单和最基本的。

分离两类的判决界由 $D_1-D_2=0$ 表示。对于任何特定的输入模式必须判定 $D_1$ 大还是 $D_2$ 大。若考虑某个函数 $D=D_1-D_2$，对于 1 类模式 $D$ 为正，对于二类模式 $D$ 为负。于是，只要处理与 $D$ 相应的一组权输入模式并判断输出符号即可进行分类。执行这种运算的分类器的原理

框图如图 9-3 所示。

在线性分类器中要找到合适的系数，以便使分类尽可能不出差错，唯一的办法就是试验法。例如，先设所有的系数为 1，送进每一个模式，如果分类有错就调整系数，这个过程就叫做线性分类器的训练或学习。例如，我们把 $N$ 个特征 $X$ 和 1 放在一起叫做 $Y$，$N+1$ 个系数为 $\omega$，即



$$Y=\begin{bmatrix}X_1\\X_2\\\vdots\\X_N\\1\end{bmatrix}\qquad \omega=\begin{bmatrix}\omega_1\\\omega_2\\\vdots\\\omega_N\\\omega_{N+1}\end{bmatrix}\tag{9.1.3}$$

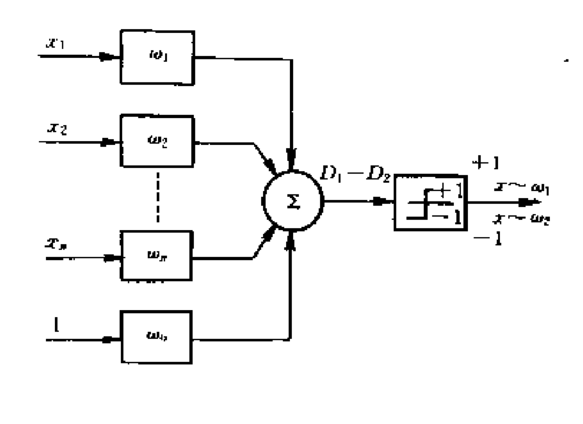



图 9-3　两类线性分类器

考虑分别属于两个不同模式类，$m=2$，此时有两个训练集 $T_1$ 和 $T_2$。两个训练集合是线性可分的，这意味着存在一个加权向量 $\omega$，若是

$$\begin{aligned}&Y^T\omega>0 \qquad && Y\in T_1\\&Y^T\omega<0 && Y\in T_2\end{aligned}\tag{9.1.4}$$

式中 $Y^T$ 是 $Y$ 的转置。如果分类器的输出不能满足式(9.1.4)的条件，可以通过“误差校正”的训练步骤对系数加以调整。例如，如果第一类模式 $Y^T\omega$ 不大于零，则说明系数不够大，可用加大系数的方法进行误差修正。具体修正方法如下：

对于任一 $Y\in T_1$，若 $Y^T\omega\leqslant 0$　　则使 $\omega^1=\omega+\alpha Y$

对于任一 $Y\in T_2$，若 $Y^T\omega>0$　　则使 $\omega^1=\omega-\alpha Y$

通常使用的误差修正方法有固定增量规则，绝对修正规则及部分修正规则。固定增量规则是选择 $\alpha$ 为一个因定的非负数。绝对修正规则是取 $\alpha$ 为一最小整数，它可使 $Y^T\omega$ 的值刚好大于零，即

$$\alpha=\text{大于}\frac{|Y^T\omega|}{Y^T\omega}\text{的最小整数}$$

部分修正规则可取 $\alpha$ 为下式所决定的值

$$\alpha=\gamma\frac{|Y^T\omega|}{Y^TY}\qquad 0<\gamma\leqslant 2\tag{9.1.5}$$

2. 最小距离分类器

线性分类器中重要的一类是用输入模式与特征空间作为模板的点之间的距离作为分类的准则。假定有 $m$ 类，给出 $m$ 个参考向量 $R_1$、$R_2$、$R_3$、…、$R_m$，$R_i$ 与模式类 $\omega_i$ 相联系。对于 $R_i$ 的最小距离分类就是把输入的新模式 $X$ 分为 $\omega_i$ 类，其分类准则就是 $X$ 与参考模型原型 $R_1$、$R_2$、$R_3$ …$R_m$ 之间的距离，跟哪一个最近就属于哪一类。$X$ 和 $R$ 之间的距离可表示为

$$|X-R_i|=\sqrt{(X-R_i)^T(X-R_i)}\tag{9.1.6}$$

其中 $(X-R_i)^T$ 是 $(X-R_i)$ 的转置，由式(9.1.6)可得

$$|X-R_i|^2=(X-R_i)^T(X-R_i)$$
$$=X^TX-X^TR_i-R_i^TX+R_i^TR_i$$
$$=X^TX-(X^TR_i+R_i^TX-R_i^TR_i)$$

由此可设定最小距离判别函数 $D_i(X)$ 为

$$D_i(X)=X^TR_i+R_i^TX-R_i^TR_i \qquad (i=1,2,3,\cdots,m) \tag{9.1.7}$$

由上边的判别函数，在分类中，如果 $X\in\omega_i$，则 $d(X,R_i)=\min$。由式(9.1.7)可见，$D_i(X)$ 是一个线性函数，因此最小距离分类器也是一个线性分类器。在最小距离分类中，在决策边界上的点与相邻两类都是等距离的，这种方法就难于解决，此时，必须寻找新的特征，重新分类。

这种分类还可以用决策区域来表示。例如有二类问题 $\omega_1,\omega_2$，其模板分别为 $R_1,R_2$，当距离 $d(X,R_1)<d(X,R_2)$，或者

$$\left[\sum_{i=1}^{n}(X_i-R_1)^2\right]^{1/2}<\left[\sum_{i=1}^{n}(X_i-R_2)^2\right]^{1/2} \tag{9.1.8}$$

则 $X\in\omega_1$，并可用决策区域来表示，如图 9-4 所示。

将模板 $R_1$、$R_2$ 作连线，再作平分线，平分线左边为 $R_1$ 区域，平分线右边为 $R_2$ 区域，$R_1R_2$ 为决策区域，中间为决策面。在这种分类中，两类情况界面为线，决策区为两平面。对于三类情况，界面为超平面，决策区为半空间。

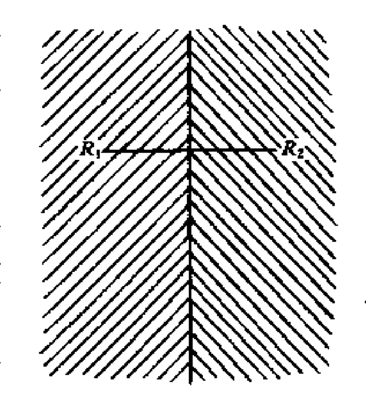


图 9-4 二类问题决策区域

3. 最近邻域分类法

最近邻域分类法是图像识别中应用较多的一种方法。在最小距离分类法中，是取一个最标准的向量作为代表。将这类问题稍微扩张一下，一类不能只取一个代表，把最小距离的概念从一个点和一个点之间的距离扩充到一个点和一组点之间的距离。这就是最近邻域分类法的基本思路。设 $R_1$、$R_2$、…、$R_m$ 分别是与类 $\omega_1$、$\omega_2$、…、$\omega_m$ 相对应的参考向量的 $m$ 个集合，在 $R_i$ 中的向量为 $R_i^k$，即 $R_i^k\in R_i$ $i=1,2,\cdots,l_i$ 也就是

$$R_i=\{R_i^1,R_i^2,\cdots,R_m^{l_i}\}$$

输入特征向量 $X$ 与 $R_i$ 之间的距离用下式表示

$$d(X,R_i)=\min|X-R_i^k| \qquad (k=1,2\cdots l_i)$$

这就是说，$X$ 和 $R_i$ 之间的距离是 $X$ 和 $R_i$ 中每一个向量的距离中的最小者。如果 $X$ 与 $R_i^K$ 之间的距离由式(9.1.6)决定，则其判决函数为

$$D_i(X)=\min\{X^TR_i^K+(R_i^K)^TX-(R_i^K)^TR_i^K\} \qquad (k=1,2,\cdots,l_i;i=1,2,\cdots,m) \tag{9.1.9}$$

设 $\qquad D_i^K(X)=X^TR_i^K+(R_i^K)^TX-(R_i^K)^TR_i^K$

则 $\qquad D_i(X)=\min\{D_i^K(X)\} \qquad (k=1,2,\cdots,l_i;i=1,2,\cdots,m)$

其中 $D_i^K(X)$ 是特征的线性组合，决策边界将是分段线性的。例如，如图 9-5 所示，有一个两类判别问题，$\omega_1$ 类的代表为 $R_1^1,R_1^2$，$\omega_2$ 类的代表为 $R_2^1,R_2^2,R_2^3$。如果有一个模式送入识别系统，首先要计算它与每个点的距离，然后找最短距离。这种方法的概念简单，分段线性边界可以代表很复杂的曲线，也可能本来是非线性边界，现在可用分段线性来近似代替。

4. 非线性判别函数

线性判别函数很简单，但也有缺点。它对于较复杂的分类往往不能胜任。在较复杂的分类问题中就要提高判别函数的次数，因此根据问题的复杂性，可将判别函数从线性推广到非线性。非线性判别函数可写成下式的形式

$$\begin{aligned} D(x) = & \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_N x_N + \\ & \omega_{12} x_1 x_2 + \omega_{13} x_1 x_3 + \cdots + \omega_{1N} x_1 x_N + \\ & \omega_{11} x_1^2 + \omega_{22} x_1^2 + \cdots + \omega_{NN} x_N^2 \\ = & \omega_0 + \sum_{k=1}^{N} \omega_{kk} x_k^2 + \sum_{k=1}^{N} \omega_k x_k + \sum_{k=1}^{N} \sum_{i=1}^{N} \omega_{ki} x_k x_i \end{aligned} \quad (9.1.10)$$

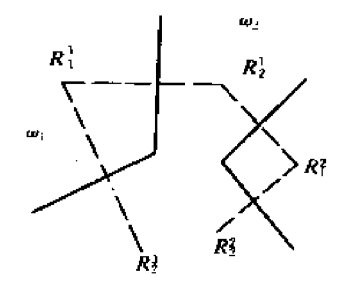


图 9-5　二类最近邻域分类

式(9.1.10)是一个二次型判别函数，通常二次型判别函数的决策边界是一个超二次曲面。

## 二、统计分类法

以上谈到的分类方法是在没有噪声干扰的情况下进行的，此时测得的特征确能代表模式。如果在抽取特征时有噪声，那么可能抽取的特征代表不了模式，这时就要用统计分类法。用统计方法对图像进行特征抽取、学习和分类是研究图像识别的主要方法之一，而统计方法的最基本内容之一是贝叶斯分析，其中包括贝叶斯决策方法、分类器、估计理论、贝叶斯学习、贝叶斯距离等等。

1. 贝叶斯公式

在古典概率中贝叶斯定理已为大家所熟习

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_{j=1}^{n} P(B_j)P(A/B_j)} \quad (9.1.11)$$

式中 $B_1$、$B_2$、…、$B_n$ 是 $n$ 个互不相容的事件，$P(B_i)$是事件 $B_i$ 的先验概率，$P(A/B_i)$是 $A$ 在 $B_i$ 已发生条件下的条件概率。贝叶斯定理说明在给定了随机事件 $B_1$、$B_2$、…、$B_n$ 的各先验概率 $P(B_i)$及条件概率 $P(A/B_i)$时，可算出事件 $A$ 出现时，去掉事件 $B_i$ 出现的后验概率 $P(B_i/A)$。假定事件 $A$ 代表肝炎病发生，而 $B_1$、$B_2$、$B_i$ 分别代表引起肝炎病发生的事件，如 $B_1$ 代表抽血时的交叉感染，$B_2$ 代表吃某种不卫生食品所引起的感染，而 $P(A/B_i)$表示在 $B_i$ 发生时，肝炎病发生的概率，则肝炎病发生时由某种原因 $B_i$ 导致的后验概率就可以用贝叶斯定理来计算。

贝叶斯公式常用于分类问题和参数估值问题中。假如设 $X$ 表示事件的状态或特征的随机变量，它可以表示图像的灰度或形状等，设 $\omega_i$ 表示事件类别的离散随机变量。对事物(比如是图像的亮度或形状)进行分类就可以用如下的公式

$$P(\omega_i/X) = \frac{P(X/\omega_i)P(\omega_i)}{\sum_{i} P(X/\omega_i)P(\omega_i)} \quad (9.1.12)$$

式中 $P(\omega_i)$称为 $\omega_i$ 的先验概率，它表示事件属于 $\omega_i$ 的预先粗略了解，$P(X/\omega_i)$表示事件属于 $\omega_i$ 类而具有 $X$ 状态的条件概率，$P(\omega_i/X)$叫做 $X$ 条件下 $\omega_i$ 的后验概率，它表示对事件 $X$ 的状态作观察后判断属于 $\omega_i$ 类的可能性。由式(9.1.12)可见，只要类别的先验概率及 $X$ 的条件概率为已知，就可以得到类别的后验概率。再加上最小误差概率或最小风险法则，就可以进行统计

判决分类。

在参数估计问题中，贝叶斯公式中二个变量常常为连续随机变量，如果写作变量 $X$ 及参数 $Q$，则有如下的公式

$$P(Q/X)=\frac{P(X/Q)P(Q)}{\int P(X/Q)P(Q)\mathrm{d}Q} \tag{9.1.13}$$

通过式(9.1.13)，由参数的先验分布 $P(Q)$ 及预先设定的条件分布 $P(X/Q)$，即可求得参数的后验分布 $P(Q/X)$。贝叶斯公式是参数估计的有力工具。

2. 贝叶斯分类法

假设有两类，每类有两种统计参数代表，即

$$\omega_1: P(\omega_1), P(X/\omega_1)$$

$$\omega_2: P(\omega_2), P(X/\omega_2) \tag{9.1.14}$$

其中 $P(\omega_1)$，$P(\omega_2)$是先验概率，$P(X/\omega_1)$、$P(X/\omega_2)$是条件概率密度函数。在噪声不确定的影响下，每个模式已不能用一个向量来表示，因此只能得到某一类模式的概率分布。

如果用贝叶斯规则的话，结果是

如果 $P(\omega_1)P(X/\omega_1)>P(\omega_2)P(X/\omega_2)$

则有 $X\in\omega_1$

如果 $P(\omega_1)P(X/\omega_1)<P(\omega_2)P(X/\omega_2)$

则有 $X\in\omega_2$

显然 $P(\omega_i)P(X/\omega_i)$ 在这里起到了判别函数的作用。在应用中，为方便起见，常取 $P(\omega_i)P(X/\omega_i)$的对数形式，即

$$\lg P(\omega_1)P(X/\omega_1) \gtrless \lg P(\omega_2)P(X/\omega_2)$$

也就是

$$\left.\begin{array}{ll}\lg\dfrac{P(X/\omega_1)}{P(X/\omega_2)}>\lg\dfrac{P(\omega_2)}{P(\omega_1)} & X\in\omega_1\\ \lg\dfrac{P(X/\omega_1)}{P(X/\omega_2)}<\lg\dfrac{P(\omega_2)}{P(\omega_1)} & X\in\omega_2\end{array}\right\}$$

在两类问题，分界面为

$$\lg P(\omega_1)P(X/\omega_1)-\lg P(\omega_2)P(X/\omega_2)=0$$

或者

$$\lg\frac{P(\omega_1)P(X/\omega_1)}{P(\omega_2)P(X/\omega_2)}=0$$

假如一个模式遵循正态分布，它的均值为$M$，协方差矩阵是$K_i$，设$m=2$，可得到其决策分界面如下，因为 $P(X/\omega_i)$ 是正态分布，所以

$$P(X/\omega_1)=(2\pi)^{-\frac{N}{2}}|K_i|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(X-M_i)^TK^{-1}(X-M_i)\right] \tag{9.1.15}$$

当 $i=1,2$ 时，按贝叶斯准则

如果
$$\lg\frac{P(X/\omega_1)}{P(X/\omega_2)}>\lg\frac{P(\omega_2)}{P(\omega_1)} \quad 则 \quad X\in\omega_1 \tag{9.1.16}$$

由式(9.1.15)和式(9.1.16)，可得到

$$-\frac{1}{2}\lg\frac{|K_1|}{|K_2|}-\frac{1}{2}[(X-M_1)^TK_1^{-1}(X-M_1)]+$$

$$\frac{1}{2}[(X-M_2)^TK^{-1}(X-M_2)]>\lg\frac{P(\omega_2)}{P(\omega_1)} \tag{9.1.17}$$

这时，两类间的决策边界是二次的。

如果两个协方差矩阵相同，即 $K_1=K_2=K$，则

$$X^TK^{-1}(M_1-M_2)+\frac{1}{2}(M_1+M_2)^TK^{-1}(M_1-M_2)>\lg\frac{P(\omega_2)}{P(\omega_1)} \quad 则有 \quad X\in\omega_1$$

$$X^{-1}K^{-1}(M_1-M_2)+\frac{1}{2}(M_1+M_2)^TK^{-1}(M_1-M_2)<\lg\frac{P(\omega_2)}{P(\omega_1)} \quad 则有 \quad X\in\omega_2$$

(9.1.18)

在这种情况下，决策边界成为线性的，所以求两类分类问题时，如果每类都是正态分布，但有不同的协方差矩阵，分界是二次函数。如果 $N$ 很大，求 $K^{-1}$相当麻烦。

除了上述方法之外，也可以用最小风险来求其类别。

考虑 $x_1,x_2,x_3,\cdots,x_N$ 是随机变量，对于每一类模式 $\omega_i,i=1,2,\cdots,m$，其 $P(X/\omega)$及 $\omega_1$ 出现的概率 $P(\omega_1)$都是已知的。以 $P(X/\omega_i)$及 $P(\omega_i)$为基础，一个分类器的成功条件是要在误识概率最小的条件下来完成分类任务。我们可定义一个决策函数 $d(x)$，其中 $d(x)=d_i$ 表示假设 $X\in\omega_i$ 被接受。如果输入模式实际是来自 $\omega_i$，而作出的决策是 $d_j$，则可用 $L(\omega_i,d_j)$表示分类器引起的损失。条件风险为

$$r(\omega_i,d)=\int L(\omega_i,d)P(X/\omega_1)\mathrm{d}X \tag{9.1.19}$$

对于给定的先验概率集 $P=\{P(\omega_1),P(\omega_2),\cdots,P(\omega_i)\}$，平均风险为

$$R(P,d)=\sum_{i=1}^{m}P(\omega_i)r(\omega_i,d) \tag{9.1.20}$$

把式(9.1.19)代入式(9.1.20)，并且令

$$r_x(P,d)=\frac{\sum_{i=1}^{m}L(\omega_i,d)P(\omega_i)P(X/\omega_i)}{P(X)} \tag{9.1.21}$$

则有 $\quad R(P,d)=\int P(X)r_x(p,d)\mathrm{d}X$

其中 $r_x(P,d)$定义为对于给定的特征向量 $X$，决策为 $d$ 的后验条件平均风险。

问题在于选择适当的决策 $d_i,i=1,2,\cdots,m$，以使平均风险 $R(p,d)$取极小，或者使条件平均风险 $r(\omega_i,d)$的极大值取极小。这种使平均风险取极小的最优决策规则称为贝叶斯规则。如果 $d^*$ 是在使平均损失极小的意义上的最优决策，则

$$r_x(p,d^*)\leqslant r_x(p,d)$$

即
$$\sum_{i=1}^{m}L(\omega_i,d^*)P(\omega_i)P(X/\omega_1)\leqslant\sum_{i=1}^{m}L(\omega_i,d)P(\omega_i)P(X/\omega_i) \tag{9.1.22}$$

对于(0,1)损失函数为

$$L(\omega_i,d_j)=\begin{cases}0 & i=j\\1 & i\neq j\end{cases}$$

平均风险实际上也就是误识的概率。在这种情形下，贝叶斯规则是

$$P(\omega_i)P(X/\omega_i)\geqslant P(\omega_j)P(X/\omega_j) \tag{9.1.23}$$

对所有的 $j=1,2,\cdots,m$。

3. 贝叶斯分类器

多类贝叶斯分类器如图 9-6 所示。其中 $P(X/\omega_i)$与 $P(\omega_i)$的乘积就是第 $i$ 类判别函数 $D_i(X)$。如果 $D_i(X)>D_j(X)$，对于一切 $i\neq j$ 的情况下，则分类器就把给定的一个特性量归于

$\omega_i$ 类。

二类贝叶斯分类器，如图 9－7 所示，在这类范畴的问题中，有时不制定二个判别函数 $D_1(X)$和 $D_2(X)$，而是定义一个判别函数。若 $D(X)>0$，则决策 $\omega_1$，否则决策 $\omega_2$。

例 9—1　已知一个判别函数为

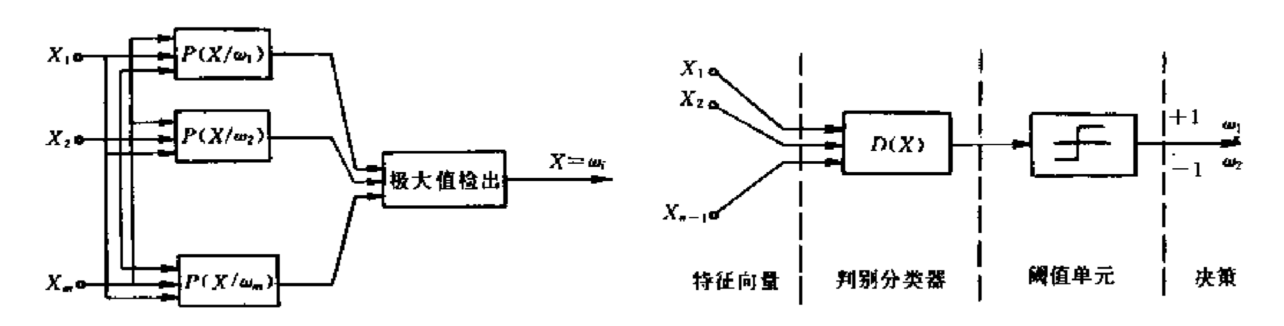



图 9－6　多类贝叶斯分类器　　　　图 9－7　二类贝叶斯分类器

$$\omega^T X=\left[1,-\frac{1}{2},-2\right]\begin{bmatrix}x_1\\x_2\\1\end{bmatrix}$$

判别规则为若 $\omega^T X>0$，则 $X$ 属于 $A$ 类；若 $\omega^T X<0$，则 $X$ 属于 $B$ 类。求 $X_1=\begin{bmatrix}7\\2\end{bmatrix}$，$X_2=\begin{bmatrix}4\\6\end{bmatrix}$，$X_3=\begin{bmatrix}6\\8\end{bmatrix}$三点各属哪一类？

解：

对 $X_1$ 点　$\left[1,-\frac{1}{2},-2\right]\begin{bmatrix}7\\2\\1\end{bmatrix}=7-1-2=4>0$，则判 $X_1$ 属 $A$ 类

对 $X_2$ 点　$\left[1,-\frac{1}{2},-2\right]\begin{bmatrix}4\\6\\1\end{bmatrix}=4-3-2=-1<0$，则判 $X_2$ 属 $B$ 类

对 $X_3$ 点　$\left[1,-\frac{1}{2},-2\right]\begin{bmatrix}6\\8\\1\end{bmatrix}=6-4-2=0$，不能判决或称为中性

以上例子的分类过程是很简单的，也是最基本的。

### 三、特征的抽取与选择

在模式识别中，确定判据是重要的，但是问题的另一方面，即如何抽取特征也是相当重要的。如果特征找不对，分类就不可能准确。特征选取的方法是很多的。从一个模式中提取什么特征，将因不同的模式而异，并且与识别的目的、方法等有直接关系。有关的方法，如图像特征的各处检测方法，曲线拟合，Hough 变换等等都在第七章中已经写到，这里不再多述。

如果把所有的特征不分主次全都罗列出来，$N$ 会很大，这也会给正确判断带来麻烦，例如图 9－8 所示。有两类模式，用两个特征 $x_1$、$x_2$ 来表达，在 $x_1$ 上的投影为 $ab$、$cd$，在 $x_2$ 上的投影的 $ef$、$gh$。那么，由图可见，$ac$ 这一段肯定属于 $\omega_1$，$bd$ 段肯定属于 $\omega_2$ 的，但是 $cb$ 段就难以分出属于哪一类，一种设想是把坐标轴作一个旋转，变成 $y_1y_2$，此时不再去测量 $x_1$、$x_2$ 而去测量 $y_1$、

$y_2$，如图 9－9 所示。由图可见，这时检测 $y_1$ 当然也分不清，可是检测 $y_2$ 就可以分得很清，这说明，当作一变换后，$y_2$ 是一个很好的特征。

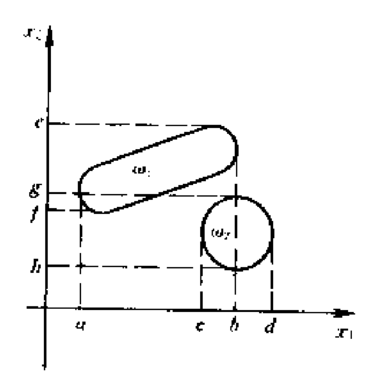


图 9－8　两类模式特征抽取之一

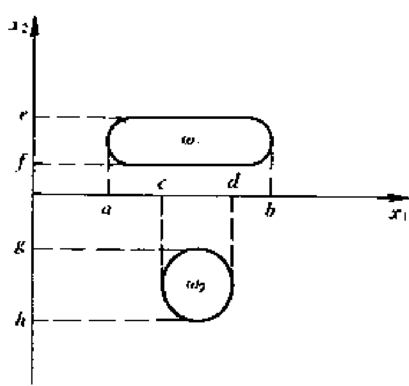


图 9－9　两类模式特征抽取之二

对于在图 9－8 和图 9－9 中所说明的特征抽取的例子中，用坐标旋转的方法得到了既少又好的特征。空间坐标的旋转就是特征空间的线性变换。究竟怎样变换才能找到较好的特征呢?其普通的方法是把每一类的协方差矩阵变成对角形矩阵，在变换后的矩阵中取其特征向量及其相对应的特征值，然后，把特征向量按其特征值的大小排列起来。特征值大的那个特征向量就是最好的特征。另外，在变换后的空间中，如果有 $m$ 个彼此关联的特征，可采用前几个最大特征值对应的特征向量作为特征，这样既可保证均方误差最小，又可大大减少特征的数目。

另外一个途径是寻找一种变换，使同一类向量靠得更近一些，以便把它聚合到一起去，在这种思想的指导下，可以找到每一类点与点之间的距离，使它最小化。这样作是应用特征值最小的那些特征向量。

假定有两类模式，测量两种特征都是正态分布，均值是 $m_1$ 和 $m_2$，这两个分布离得越远越容易识别。所谓离得远不一定是均值相差较远。在这种情况下，不能用点与点之间的距离，也不是点与一组点间的距离，而是两个分布间的距离，这是一个统计距离。如果在统计意义上两类离得远就容易识别。如果有 $M$ 个特征，就要计算它们的统计距离，哪个特征上统计距离最远，哪个特征就最好。一般计算统计距离的方法有许多。例如：贝叶斯误差概率，疑义度或仙农熵，贝叶斯距离，广义柯尔莫哥洛夫距离等等。

## §9.2　遥感图像的几种监督分类方法

有了以上的基本知识，我们结合遥感图像介绍具体的分类方法。遥感图像分类有两大类：监督分类和非监督分类。

监督分类也称训练场地法，根据先验的类别知识，选择有代表性的场地作试验样本特征参数的实际观测，如测其光谱特性参数，并求其统计特征如均值 $\mu_i$、协方差矩阵 $\Sigma_j$ 和离散度等等，由此初步建立一个判别函数，用来判别分类其它未知类别的样品，若分类精度不够，再修改调整或重建，直到分类精度满足要求为止。这样就可以应用最后的判别函数来作为正式分类的设计依据。这个过程就是有导师的学习过程，即监督学习，也就是监督分类。

监督分类的原理框图如图 9－10 所示，监督分类方法有许多种，如判别分析、序贯分析、最大似然分析和监督分类的图形识别等等，下面选几个扼要介绍一个。

## 一、判别分析

判别分析适用于变量呈正态分布的情况下。如在卫星照片的地质分类中常被应用。其基本思路是将高维数矢量空间投影(变换)到低维矢量空间,从而使不同类别的目标物体特征参数尽可能地分开,便于判决分类的实现。从一个简单的例子开始,如图 9-11 中有两类物体 $A$ 和 $B$,由于两类物体在 $x_1$ 轴和 $x_2$ 轴上分布有重叠,直接使用其中一个变量 $x_1$ 或 $x_2$ 是无法分开的,但如果将 $x_1$ 和 $x_2$ 变换到一个变量 $R$ 上,使两类物体分布在 $R$ 轴上投影中心($R_A$ 和 $R_B$)距离尽可能近,那么就可以选择两个投影中心的中点 $R_0$ 作为判别点。其投影在 $R_0$ 和 $R_A$ 之间的任何 $X(x_1,x_2)$ 均判别为 $A$ 类。在 $R_0$～$R_B$ 之间的判别为 $B$ 类。

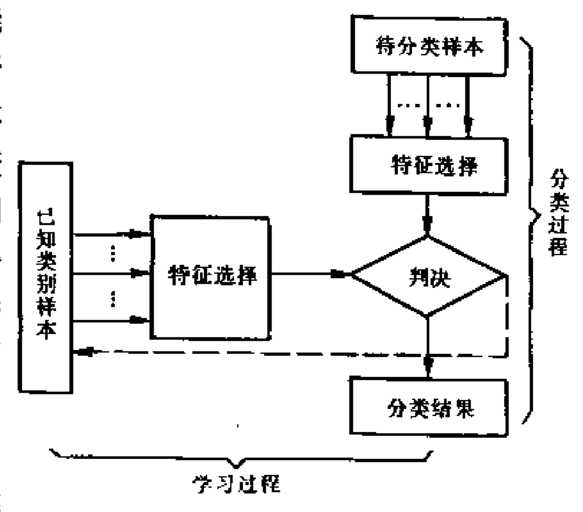
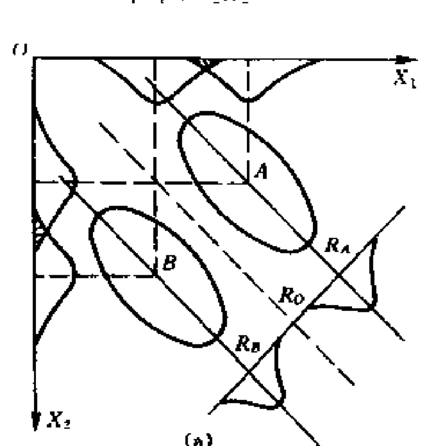


图 9-10 监督分类框图

假定上述变换是线性变换,即 $R$ 是 $x_1$ 和 $x_2$ 的线性函数,且表示为

$$R=\lambda_1 x_1+\lambda_2 x_2$$



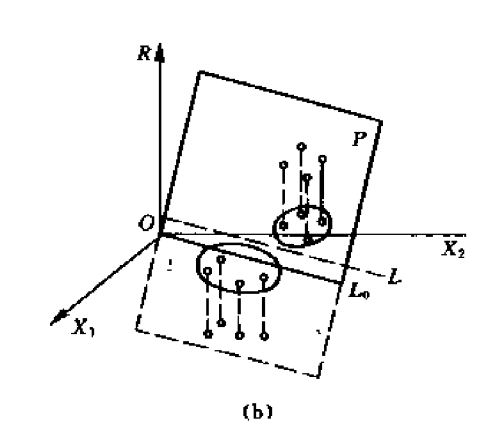




图 9-11 二维变量的判别分析

(a)二维到一维变量变换;(b)判别函数的几何意义

那么应用几何学知识就可以理解。根据 $\lambda_1$ 和 $\lambda_2$ 作一个平面 $P$,使 $P$ 与 $x_1-x_2$ 平面相交 $L$,从 $x_1-x_2$ 面内物体 $A$ 和 $B$ 上的点向 $P$ 平面作垂直线且和 $x_1-x_2$ 平面相交,这些交点到 $x_1-x_2$ 平面的距离即为物体 $A$ 和 $B$ 上点对应的 $R$ 值。由于 $P$ 平面的走向和倾斜度是由 $\lambda_1$ 和 $\lambda_2$ 决定的,因此可以通过选择合适的 $\lambda_1$ 和 $\lambda_2$ 使物体 $A$ 对应的 $R$ 值尽可能"紧凑",物体 $B$ 对应的 $R$ 值也尽可能"紧凑"。而使它们之间的距离尽可能的远。这样,就可以通过对未知点 $R$ 值的判别来分类,因此 $\lambda_1$ 和 $\lambda_2$ 值的选择是关键。设 $A$、$B$ 二类物体所有点的总期望值点所对应的 $R$ 值为 $R_0$,$A$、$B$ 类各自内部点期望值点所对应的 $R$ 值分别为 $R_A$ 和 $R_B$,$n_1$ 和 $n_2$ 分别为 $A$、$B$ 类的点数。为了使类内 $R$ 值"紧凑",即使其类内 $R$ 值方差最小,而又希望类间 $R$ 值相差大,即使其类间方差为最大,可通过下式运算

$$G=n_1(R_A-R_0)^2+n_2(R_B-R_0)^2 \tag{9.2.1}$$

使其趋于最大，而下式

$$H=\frac{1}{n_1+n_2-2}\Big[\sum_{j=1}^{n_1}(R_j-R_A)^2+\sum_{j=n_1+1}^{n_1+n_2}(R_j-R_B)^2\Big]$$

趋于最小。式中

$A$ 类均值：
$$R_A=\frac{1}{n_1}\sum_{j=1}^{n_1}(\lambda_1 x_{1j}^A+\lambda_2 x_{2j}^A)=\lambda_1\overline{X}_1^A+\lambda_2\overline{X}_2^A$$

$B$ 类均值：
$$R_B=\frac{1}{n_2}\sum_{j=n+1}^{n_1+n_2}(\lambda_1 x_{1j}^B+\lambda_2 x_{2j}^B)=\lambda_1\overline{X}_1^B+\lambda_2\overline{X}_2^B$$

两类均值：
$$R_0=\frac{1}{n_1+n_2}\sum_{j=1}^{n_1+n_2}(\lambda_1 x_{1j}+\lambda_2 x_{2j})=\lambda_1\overline{X}_{10}+\lambda_2\overline{X}_{20}$$

式中：$\overline{X}_1^A=\frac{1}{n_1}\sum_{j=1}^{n_1}X_{1j}^A$ $\qquad$ $\overline{X}_2^A=\frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2}X_{1j}^B$

$\overline{X}_1^B=\frac{1}{n_1}\sum_{j=1}^{n_1}X_{2j}^A$ $\qquad$ $\overline{X}_2^B=\frac{1}{n_2}\sum_{j=n_2+1}^{n_1+n_2}X_{2j}^B$

$\overline{X}_{10}=\frac{1}{n_1+n_2}\sum_{j=1}^{n_1+n_2}X_{1j}$ $\qquad$ $\overline{X}_{20}=\frac{1}{n_1+n_2}\sum_{j=1}^{n_1+n_2}X_{2j}$

将上述各式代入式(9.2.1)，即可获得

$$G=\frac{n_1n_2}{n_1+n_2}[\lambda_1(\overline{X}_1^A-\overline{X}_1^B)+\lambda_2(\overline{X}_2^A-\overline{X}_2^B)]^2 \tag{9.2.2}$$

$$H=\lambda_1^2S_{11}+2\lambda_1\lambda_2S_{12}+\lambda_2^2S_{22} \tag{9.2.3}$$

式中 $\qquad S_{11}=\frac{1}{n_1+n_2-2}\Big[\sum_{j=1}^{n_1}(X_{ij}^A-\overline{X}_1^A)^2+\sum_{j=n_1+1}^{n_1+n_2}(X_{ij}^B-\overline{X}_1^B)^2\Big]$

为 $X_1$ 的总方差；

$$S_{22}=\frac{1}{n_1+n_2-2}\Big[\sum_{j=1}^{n_1}(X_{2j}^A-\overline{X}_2^A)^2+\sum_{j=n_1}^{n_1+n_2}(X_{2j}^B-\overline{X}_2^B)^2\Big]$$

为 $X_2$ 的总方差；

$$S_{12}=\frac{1}{n_1+n_2-2}\Big[\sum_{j=1}^{n_1}(X_{1j}^A-\overline{X}_1^A)(X_{2j}^A-\overline{X}_2^A)+\sum_{j=n_1+1}^{n_1+n_2}(X_{1j}^B-\overline{X}_1^B)(X_{2j}^B-\overline{X}_2^B)\Big] \tag{9.2.4}$$

为 $X_1$ 和 $X_2$ 的总方差。

为使 $G$ 最大，$H$ 最小同时满足，即可要求 $G/H=T^2$ 为最大，两边取对数

$$\ln T^2=\ln G-\ln H \tag{9.2.5}$$

将式(9.2.5)分别对 $\lambda_1$ 和 $\lambda_2$ 求偏导数并令其为零，则得到

$$\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}=\begin{bmatrix}\overline{X}_1^A-\overline{X}_1^B\\ \overline{X}_2^A-\overline{X}_2^B\end{bmatrix}$$

解此方程即可得到 $\lambda_1$ 和 $\lambda_2$ 为

$$\lambda_1=S_{11}(\overline{X}_1^A-\overline{X}_1^B)-S_{12}(\overline{X}_2^A-\overline{X}_2^B)/S_{11}S_{22}-S_{12}^2$$
$$\lambda_2=S_{11}(\overline{X}_2^A-\overline{X}_2^B)-S_{12}(\overline{X}_1^A-\overline{X}_1^B)/S_{11}S_{22}-S_{12}^2 \tag{9.2.6}$$

以 $A$ 类和 $B$ 类所有点结合在一起的总期望值点坐标 $X_{10}$,$X_{20}$代入式

$$R=\lambda_1 x_1+\lambda_2 x_2 \tag{9.2.7}$$

中可以得到

$$R_0=\lambda_1 X_{10}+\lambda_2 X_{20} \quad 或 \quad \lambda_1 X_{10}+\lambda_2 X_{20}-R_0=0 \tag{9.2.8}$$

通常称式(9.2.8)为判别直线,如图 9-11 中的直线 $L$,称 $R_0$ 为判别指数,而式(9.2.7)一般称之为判断函数或判别方程。以上讨论的是二维的情况,二类多维的情况依次类推,其判决函数为

$$R=\lambda_1 X_1+\lambda_2 X_2+\cdots+\lambda_K X_K$$

式中,$\lambda_1,\lambda_2,\cdots,\lambda_K$ 由下列方程组解得

$$\begin{bmatrix} S_{11} & \cdots & S_{1K} \\ \vdots & & \\ \vdots & & \\ S_{K1} & \cdots & S_{KK} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{bmatrix} = \begin{bmatrix} \overline{X}_1^A-\overline{X}_1^B \\ \overline{X}_2^A-\overline{X}_2^B \\ \vdots \\ \overline{X}_K^A-\overline{X}_K^B \end{bmatrix}$$

简化为　　$\Sigma\lambda=D$　或　$\lambda=\Sigma^{-1}D$

其中 $\Sigma$ 称为 $X_1X_2\cdots X_K$ 的协方差矩阵;$D$ 称为均值差向量之列矩阵;$\lambda$ 称为判决函数中系数 $\lambda$ 之列矩阵。

例 9-2　建立区分海滩沙和滨外沙的判别函数,并判别一个未知样本 $X_1=0.33$,$X_2=1.17$属于哪一类?

解:(1)确定训练场地:

①特征参数选择是根据需要和可能,应根据服务对象学科的知识和经验来确定选用样本的什么参数对分类最为合适。本例根据地学知识用“中等颗粒大小”和“分选系数”作为两个特征,分别以 $X_1$ 和 $X_2$ 表示。

②特征参数的测量,对 $X_1$ 和 $X_2$ 测量次数分别表示为 $n_1$ 和 $n_2$。$n_1$ 和 $n_2$ 可以相等,也可以不等,但次数应足够多,能在统计意义上有效,例如实际结果为表 9-1。

表 9-1　海沙样本测量结果

| A 类(海滩沙) | | B 类(滨外沙) | |
|---|---|---|---|
| $X_1^A$ | $X_2^A$ | $X_1^B$ | $X_2^B$ |
| 中等颗粒大小 | 分选系数 | 中等颗粒大小 | 分选系数 |
| 0.333 | 1.08 | 0.339 | 1.12 |
| 0.340 | 1.08 | 0.336 | 1.12 |
| 0.338 | 1.09 | 0.350 | 1.12 |
| 0.333 | 1.10 | 0.352 | 1.13 |
| ⋮ | ⋮ | ⋮ | ⋮ |

③求各项均值

$$\overline{X}_1^A=\frac{1}{n_1}\sum_{j=1}^{n_1}X_{1j}^A$$
$$=\frac{1}{34}(0.333+0.340+\cdots+0.326+\cdots+0.330)$$

$$= 0.330$$

$$\overline{X}_2^A = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{2j}^A$$

$$= \frac{1}{34}(1.08 + 1.08 + \cdots + 1.18 + \cdots + 1.17)$$

$$= 1.167$$

$$\overline{X}_1^B = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{1j}^B$$

$$= \frac{1}{47}(0.339 + 0.336 + \cdots + 0.342 + \cdots + 0.340) = 0.340$$

$$\overline{X}_2^B = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}^B$$

$$= \frac{1}{47}(1.12 + 1.12 + \cdots + 1.20 + \cdots + 1.21) = 1.210$$

(2)求均值差向量 $D$:

$$D=\begin{bmatrix} D_1 \\ D_2 \end{bmatrix}=\begin{bmatrix} \overline{X}_1^A-\overline{X}_1^B \\ \overline{X}_2^A-\overline{X}_2^B \end{bmatrix}=\begin{bmatrix} 0.330-0.340 \\ 1.167-1.210 \end{bmatrix}=\begin{bmatrix} -0.010 \\ -0.043 \end{bmatrix}$$

(3)求协方差矩阵 $\sum$:

$$\sum=\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}=\begin{bmatrix} 0.00003 & -0.00017 \\ -0.00017 & 0.00231 \end{bmatrix}$$

其中 $S_{11}$、$S_{12}$、$S_{22}$等由式(8.6.4)求得

$$S_{11} = \frac{1}{n_1 + n_2 - 2}\left[\sum_{j=1}^{n_1}(X_{1j}^A - X_1^A)^2 + \sum_{j=n_1+1}^{n_1+n_2}(X_{1j}^B - X_1^B)^2\right]$$

$$= 0.00003$$

$$S_{12}=S_{21}=-0.00017 \qquad S_{22}=0.00231$$

(4)根据矩阵的运算法求出逆矩阵$\sum^{-1}$:

$$\sum^{-1} = \frac{1}{|\sum|}\begin{bmatrix} S_{22} & -S_{12} \\ -S_{21} & S_{11} \end{bmatrix}=\begin{bmatrix} 57178.217 & 4207.912 \\ 4207.912 & 742.574 \end{bmatrix}$$

其中 $|\sum| = \begin{vmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}$

(5) 再由 $\lambda = \sum^{-1} D$ 式求出判别函数 $\lambda$:

$$\lambda = \sum^{-1} D = \begin{bmatrix} 57178.217 & 4207.912 \\ 4207.912 & 742.574 \end{bmatrix}\begin{bmatrix} -0.01 \\ -0.043 \end{bmatrix}$$

$$= \begin{bmatrix} -752.72 \\ -74.01 \end{bmatrix}$$

即:$\lambda_1=-752.72$;$\lambda_2=-74.01$

(6)由式(9.2.8)确定判别函数:

$$R=\lambda_1 X_1+\lambda_2 X_2=-752.72X_1-74.01X_2$$

(7)最后求出 $R_A$、$R_B$ 和 $R_0$ 为

$$R_A=\lambda_1\overline{X}_1^A+\lambda_2\overline{X}_2^A$$

$$=-752.72\times0.330+(-74.01\times1.167)=-334.77$$

$$R_B=\lambda_1\overline{X}_1^B+\lambda_2\overline{X}_2^B$$

$$=-752.72\times0.340+(-74.01\times1.210)=-345.48$$

$$R_0=\frac{1}{n_1+n_2}(n_1R_A+n_2R_B)$$

$$=\frac{1}{34+47}[34\times(-334.77)+47(-345.48)]$$

$$=-340.98$$

根据上述求得的 $A$ 类和 $B$ 类中心判别值 $R_A$ 和 $R_B$ 以及判别指数 $R_0$，可画出其直方图如图 9-12 所示。

(8)判别未知样本 $X_1=0.33, X_2=1.77$ 属哪一类？

$$R=\lambda_1X_1+\lambda_2X_2$$

$$=-752.72\times0.33+(-74.01\times1.17)=-334.99$$

因为未知数样本的 $R$ 值在 $R_0$ 和 $R_A$ 之间，故判此类样本为 $A$ 类，为海滨沙。

在实际问题中还有一些问题可以说明一下：

①若计算出 $R_0$ 过于偏向 $R_A$ 或 $R_B$，可调整一下两类训练样本取数 $n_1$ 和 $n_2$。

②对未知数样本的 $R$ 值到 $R_A$ 或 $R_B$ 之距离不要太大，一般情况下，应满足

$$\left.\begin{matrix}|R-R_A|\\|R-R_B|\end{matrix}\right\}<100$$

若超过 100 往往不能判别为 $A$ 类或 $B$ 类。

$R_A=-334.77$　$R_0=-340.98$　$R_B=-345.48$

图 9-12　$A$ 类和 $B$ 类的直方图

## 二、序贯分析

序贯分析法实际是就是判别分析法的多次应用。其过程可用树状图来表示，图 9-13 给出两种常见的树状图。图中，$DP_1$、$DP_2$、$DP_3$、$DP_4$ 为四个分类点，每个分类点处有一个判别函数与之对应。按分类点次序逐个应用对应的判别函数，即可将原始图像样本分为多类。这里我们不再进行系统的理论分析，只结合一个具体的例子来说明一下分类过程，虽有其局限性，但仍不失一般性。

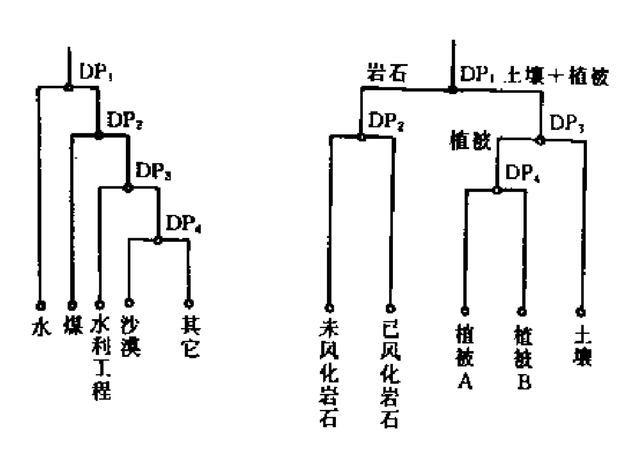



图 9-13　两种树状图

例 9-3　已知①训练组各分类点处各波段的亮度值如表 9-2 所示。②各分类点见(9-13)图处的判别函数分别为

$$R_{DP1}=32.38\times ch7/6-22.62\times ch7/4+5.52\times ch6/5+16.54\times ch6/4$$

$$R_{DP2}=-66.22\times ch7/6-30.82\times ch6/4$$

$$R_{DP3}=-89.47\times ch7/6+86.69\times ch7/5+1.0\times ch7/4-63.01\times ch6/5$$

$R_{DP_4}=-61.52\times ch7/6-29.14\times ch6/4$

试对表 9-3 所给的十个象素进行分类。

表 9-2

| | $DP_1$ | | $DP_2$ | | $DP_3$ | | $DP_4$ | |
|---|---|---|---|---|---|---|---|---|
| | 水的 $\overline{X}$ | 其它类 $\overline{X}$ | 煤的 $\overline{X}$ | 其它类的 $\overline{X}$ | 水利工程 $\overline{X}$ | 其它类 $\overline{X}$ | 沙漠 $\overline{X}$ | 其它类 $\overline{X}$ |
| $ch4$ | 11.05 | 29.45 | 11.05 | 29.72 | 25.79 | 29.97 | 34.02 | 27.28 |
| $ch5$ | 7.98 | 43.04 | 10.52 | 43.53 | 37.01 | 43.93 | 52.84 | 38.03 |
| $ch6$ | 2.52 | 46.35 | 9.15 | 46.87 | 56.66 | 46.26 | 56.57 | 39.43 |
| $ch7$ | 0.05 | 36.74 | 3.30 | 37.23 | 52.19 | 36.30 | 44.57 | 30.56 |

表 9-3

| | $ch7/6$ | $ch7/5$ | $ch7/4$ | $ch6/5$ | $ch6/4$ | $ch5/4$ |
|---|---|---|---|---|---|---|
| 1 | 0.8973 | 1.2420 | 1.929 | 1.384 | 2.150 | 1.553 |
| 2 | 0.9018 | 1.0260 | 1.430 | 1.138 | 1.586 | 1.393 |
| 3 | 0.8902 | 0.9059 | 1.262 | 1.017 | 1.418 | 1.393 |
| 4 | 0.0118 | 0.0061 | 0.0041 | 0.5233 | 0.3474 | 0.0638 |
| 5 | 0.7312 | 0.7419 | 1.203 | 1.111 | 1.646 | 1.622 |
| 6 | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

**解:**(1)先计算两个分类中心 $R_A$ 和 $R_B$ 以及判别指数 $R_0$

①计算 $DP_1$ 处的 $R_{水}$ 和 $R_{其它}$及 $R_0$

$$R_{水}=32.38\times\frac{0.05}{2.52}-22.62\times\frac{0.05}{11.05}+5.52\frac{2.52}{7.98}+16.54\times\frac{2.52}{11.06}=6.04$$

$$R_{其它}=32.38\times\frac{36.74}{46.33}-22.62\frac{36.74}{29.45}+5.52\frac{46.33}{43.04}+16.54\times\frac{46.33}{29.45}=29.42$$

$$R_0=\frac{1}{2}(R_{水}+R_{其它})=\frac{1}{2}(6.04+29.42)=17.73$$

对 $DP_2$、$DP_3$ 和 $DP_4$ 以同样的方法计算,其结果如下

$$②DP_2=\begin{cases}R_{煤}=-47.60\\R_{其它}=-97.23\\R_0=-73\end{cases}\quad ③DP_3=\begin{cases}R_{水利}=-54\\R_{其它}=-63.8\\R_0=-58.9\end{cases}\quad ④DP_4=\begin{cases}R_{沙漠}=-96.92\\R_{其它}=-89.80\\R_0=-93.36\end{cases}$$

(2)将 10 个象素的比值变量(表 9-3)依 $DP_1$、$DP_2$、…、$DP_3$ 和 $DP_4$ 次序代入各处判决函数 $R_{DP1}$、$R_{DP2}$、$R_{DP3}$,若能判别出属于哪一类,则结束这个象素的运算。否则再依次向下运算,直到求出该象素属于何类为止。例如对第一个象素运算过程如下:

①先用 $R_{DP1}$计算

$$R_1=32.38\times0.8973-22.62\times1.929+5.52\times1.384+16.54\times2.150=28.6$$

根据上步计算可见

$$R_0=17.73<R_1'=28.6<R_{其它}=29.42$$

故判第一象素属"其它类"。必须继续运算,用 $DP_2$ 判别函数进行运算

$$R_1=-61.22\times0.8973-30.82\times2.150=-121.20$$

因为 $R_1=-121.20<R_{其它}=-97.23<R_0=-73$

可见第 1 象素也不是煤,还要继续判别,将第一象素比值变量代入 $DP_3$ 处判别函数进行计算

$$R_1=-89.47\times0.8973+86.69\times1.242+1.0\times1.929-63.1\times1.384=-57.89$$

可见 $R_0=-58.9<R_1=-57.89<R_{水利工程}=54$

因此可以判决第一象素是属于水利工程类。

(3)按同样的方法计算其它 9 个象素所属类别见表 9-4。

为了改进分类速度，往往按变量分类能力来排定分类次序，而变量分类能力以 $F$ 值表示，$F$ 增大，分类能力强，最先使用分类，建立判别函数。

表 9-4

| 象素号 | $DP_1$ 处 $R$ | | | $DP_2$ 处 $R$ | | | $DP_3$ 处 $R$ | | | $DP_4$ 处 $R$ | | | 判别结果 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $R_{水}$ | $R_0$ | $R_{其它}$ | $R_0$ | $R_{森}$ | $R_{其它}$ | $R_0$ | $R_{水利}$ | $R_{沙}$ | $R_0$ | $R_{其它}$ | $R_{沙}$ | |
| | 6.04 | 17.73 | 29.42 | -73 | -47.6 | -97.23 | -58.9 | -54 | -6.38 | -93.36 | -89.8 | -96.92 | |
| 1 | 28.6 | | | −121.20 | | | −57.89 | | | | | | 水利工程 |
| 2 | 19.5 | | | −104.0 | | | −52.02 | | | −101.69 | | | 沙漠 |
| 3 | 29.4 | | | −98.3 | | | −77.2 | | | −96.3 | | | 沙漠 |
| 4 | 3.54 | | | | | | | | | | | | 水 |
| 5 | 29.4 | | | −95.5 | | | −64.7 | | | −93 | | | 其它 |
| 6 | ⋮ | | | ⋮ | | | ⋮ | | | ⋮ | | | ⋮ |

## 三、似然比判别分析

在某些情况下，例如遥感图像中的一些类的物质波谱特性，虽经特征提取(就是相当于投影到 $R$ 轴上的过程)，但不能应用一根直线即线性判别函数把它们分开，必须建立一种非线性判别函数来分类。似然比判别分析方法的主要概念是：用待分类的象点特征参数(灰度值，波段比值等)构成特征矢量 $X$。如果其条件概率密度 $P(X/\omega_i)$和$P(X/\omega_j)$以及先验概率 $P(\omega_i)$和 $P(\omega_j)$是已知的话，那么就可以用最小错误概率的贝叶斯判决方法，也就是最大似然判决。其判别准则可写为

$$L(X)=\frac{P(X/\omega_i)}{P(X/\omega_j)}\gtrless\frac{P(\omega_j)}{P(\omega_i)}=X\in\begin{cases}\omega_i\\ \omega_j\end{cases} \tag{9.2.9}$$

式中 $P(\omega_j)/P(\omega_i)$往往称为似然比判别分析阈值。

为了运算方便，上式可用对数形式表示为

$$-\ln L(X)=-\ln P(X/\omega_i)+\ln P(X/\omega_j)\lesseqgtr\ln\frac{P(\omega_j)}{P(\omega_i)}\Diamond X\in\begin{cases}\omega_i\\ \omega_j\end{cases} \tag{9.2.10}$$

当条件概率密度 $P(X/\omega_K)$都是期望矢量 $M_K$，协方差矩阵为 $\Sigma_K$ 的正态分布的时候，变量 $X$ 的维数为 $n$，则

$$P(X/\omega_i)=(2\pi)^{-\frac{n}{2}}\left|\sum\nolimits_i\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(X-M_i)^T\sum\nolimits_i^{-1}(X-M_i)\right] \tag{9.2.11}$$

$$P(X/\omega_j)=(2\pi)^{-\frac{n}{2}}\left|\sum\nolimits_i\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(X-M_j)^T\sum\nolimits_i^{-1}(X-M_j)\right] \tag{9.2.12}$$

式(9.2.10)变成

$$h(X)=-\ln L(X)=\frac{1}{2}(X-M_i)^T\sum\nolimits_i^{-1}(X-Mi)-\frac{1}{2}(X-M_2)^T\sum\nolimits_j^{-1}(X-M_j)+\frac{1}{2}\ln\frac{\left|\sum i\right|}{\left|\sum j\right|}\lesseqgtr\ln\frac{P(\omega_i)}{P(\omega_j)}\Diamond X\in\begin{cases}\omega_i\\ \omega_j\end{cases} \tag{9.2.13}$$

由式(9.2.13)可见，这个似然比判别函数是 $X$ 的二次型函数。如果 $n=2$，那么判别边界就是一条曲线，见图 9-14(a)。

如果还满足 $\Sigma_1=\Sigma_2=\Sigma$ 时，则式(9.2.13)变为

$$h(X)=(M_2-M_1)^T\Sigma^{-1}X+\frac{1}{2}(M_1^T\Sigma^{-1}M_1-M_2^T\Sigma^{-1}M_2)\lesseqgtr\ln\frac{P(\omega_1)}{P(\omega_2)}\Rightarrow X\in\begin{cases}\omega_1\\\omega_2\end{cases} \tag{9.2.14}$$

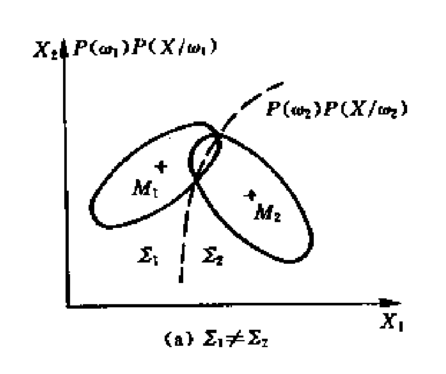
如果采用白化变换，使 $\Sigma_1=\Sigma_2=\Sigma=I$（$I$ 为单位矩阵）时，则式(9.2.14)还可简化为

$$(X-M_1)^T(X-M_1)-(X-M_2)^T(X-M_2)\lesseqgtr 2\ln\frac{P(\omega_1)}{P(\omega_2)}\Rightarrow X\in\begin{cases}\omega_1\\\omega_2\end{cases} \tag{9.2.15}$$

或 $$\|X-M_1\|^2-\|X-M_2\|^2\lesseqgtr 2\ln\frac{P(\omega_1)}{P(\omega_2)}\Rightarrow X\in\begin{cases}\omega_1\\\omega_2\end{cases}$$

式中 $$\|X-M_1\|^2=\sum_{i=1}^{n}(Xi-M_{1i})^2 \tag{9.2.16}$$

$$\|X-M_2\|^2=\sum_{i=1}^{n}(X_i-M_{2i})^2 \tag{9.2.17}$$



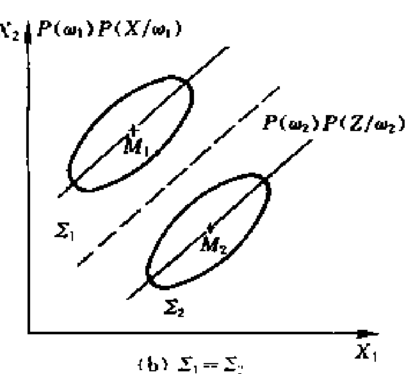




图 9－14　两个正态分布的判决边界

式(9.2.16)和式(9.2.17)分别表示 $X$ 到 $M_1$ 和 $M_2$ 距离的平方，$n$ 为 $X$ 的维数，$X_i$ 为 $X$ 的第 $i$ 个分量，$M_{1i}$和 $M_{2i}$分别表示 $M_1$ 和 $M_2$ 中的第 $i$ 个分量，$2\ln\frac{P(\omega_1)}{P(\omega_2)}$为阈值。由此可见，式(9.2.15)的含义是：以阈值 $2\ln\frac{P(\omega_1)}{P(\omega_2)}$为标准，比较 $X$ 到 $M_1$ 和 $M_2$ 的距离，当两个距离大于上述阈值标准时，判 $X$ 属于 $\omega_1$ 类，反之则判 $X$ 属于 $\omega_2$ 类。如果出现 $P(\omega_1)=P(\omega_2)$，即阈值标准为0.5时，则判别边界最简单，为 $M_1$ 和 $M_2$ 连线的垂直平分线。以上只简单介绍了几种，关于分类识别的方法还很多，这里不能一一加以介绍了。

## 四、用傅里叶描绘子实时识别图像

下面将简单介绍由 TMS 320 C25 的处理芯片所完成的实时图像识别的装置，它的硬件具有两块实时图像数据存贮区，使图像采集、处理、显示同时进行。配上 A/D 和 D/A 板，就可以构成一个完整的图像目标实时识别系统。该系统由 TMS 320 C25 完成整机控制，它的特点是不依赖于 PC 机而独立工作。它的软件包括目标轮廓提取、目标形心提取、傅里叶描绘子和匹配识别等过程。

### 1. 图像目标的轮廓提取

在目标与背景之间有一定差别的情况下，如果知道了图像目标的外形轮廓，也就确定了该目标，可以采取轮廓跟踪方法提取目标轮廓。首先对图像进行光栅扫描搜索，寻找满足条件（灰度大于某个阈值）的象素作为起始点 $S_0$，从 $S_0$ 开始逆时针探讨 $S_0$ 的 8 邻域象素，找到满足条

件的第二个轮廓点 $S_1$，依次继续寻找其它的轮廓点，此时，$S_0$、$S_1$、$S_2$、…、$S_{m-1}$形成了该图像目标的封闭轮廓。

2. 基于傅里叶描绘子的实时识别

当图像目标轮廓上的点一旦被确定之后，就可以直接从这些轮廓点中提取信息，产生一组描绘子作为图像目标的特征。假如我们已取好了目标的形心(图 9-15)，轮廓点对 $X$ 轴形成角度为 $Q_i$，形心到轮廓点的距离为 $r$。可以说，序列 $\{r(Q_i)\}$ 很好地反映了图像目标轮廓的形状特征，若以 $Q$ 为角度增量，等角度地求形心到轮廓的距离 $r(i)$，则距离序列可以写作：$r=\{r(0),r(1),r(2),\cdots,r(N-1)\}$，$N=2\pi/Q$ 对序列 $\{r(i)\}$ 进行傅里叶变换得

$$R(n) = A(n) + \mathrm{j}B(n) = \sum_{i=0}^{N-1} r(i)\exp[-\mathrm{j}2\pi ni/N]$$

$$(n = 0,1,2,\cdots,N-1)$$

这里的 $R(n)$ 就形成了目标轮廓的特征量。

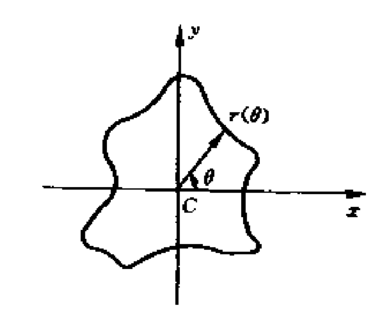


图 9-15 从形心学直线到轮廓的距离

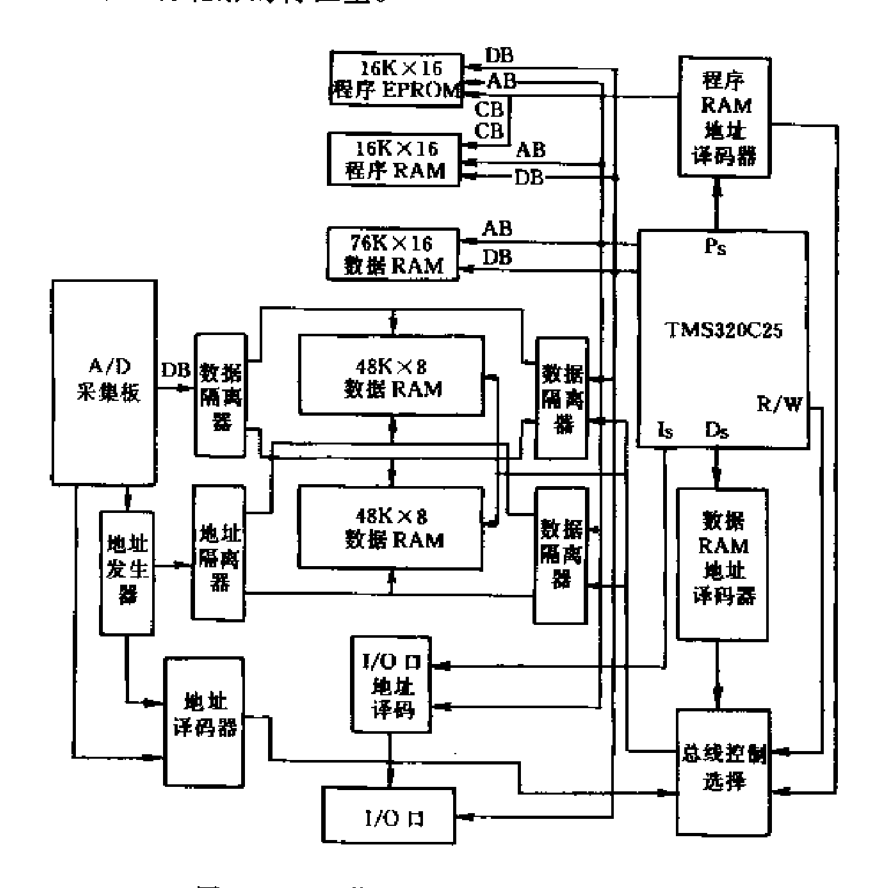


图 9-16 图像的实时识别装置的原理框图

若目标在运动中不发生旋转变化或旋转角度较小，我们可以用下面的方法对目标进行识别。

设 $r=\{r(0),r(1)\cdots r(N-1)\}$ 是标准目标的傅里叶描绘子。$r'=\{r'(0),r(1)\cdots r'(N-1)\}$ 是实时图像中待识目标的傅里叶描绘子，我们用它们对应分量的绝对差之和作为识别准则，$T$ 为判决阈值则。

$$d = \sum_{i=0}^{N-1} |r(i) - r'(i)|$$

若 $d \geq T$ 识别失败，不是目标；$d < T$，识别成功，是目标。

由于实际目标常常发生旋转，则要对上述方法进行改进，有许多种方法，我们设法能在目标上找到一个基准轴，这个轴相对目标应当是不变的，从该基准轴等角度地求形心到轮廓的距离，便得到一组距离序列，这时就可以运用上面的绝对差之和的准则进行识别。由于目标的主轴计算涉及到目标的一二阶矩计算，比较麻烦，我们就采用下面的近似选取基准轴的方法。

以上共提取了 $N$ 个形心到轮廓的距离，它相当于 $N/2$ 个过形心的轴，以 $N/2$ 个轴中最长的那个轴作为基准轴，同样也能避免目标旋转对识别的影响。在实验中，$N$ 取的越大，效果愈好。

前面谈到形心的求取，可以在轮廓点集 $\{(x_i, y_i)\ i = 1, 2, \cdots, M\}$ 求得之后，求取其形心 $0(x_0, y_0)$

$$X_0 = \frac{1}{M}\sum_{i=1}^{M} x_i \qquad Y_0 = \frac{1}{M}\sum_{i=1}^{M} y_i$$

3. 实时识别装置的原理框图

它是一个不依赖 PC 而独立工作的装置，主要包括程序区、数据区、输入输出口、控制电路等等，以上电路由于篇幅有限，不再详细介绍。

## §9.3 模糊模式识别

### 一、综述

“模糊”一词，译自英文“Fuzzy”，意为“模糊的”，“不分明的”。1965 年美国控制论专家 L. A. Zadch 首先将“Fuzzy”一词引入数学界，他在“Information and Control”杂志 1965 年第 8 期中发表了“Fuzzy Sets”一文，标志着模糊数学的诞生。

作为控制论专家的 Zadeh，工作性质使他多年来战斗在精确性与模糊性搏斗的战场。他认为现有的数学大多数是根据力学、物理学、天文学的发展而建立起来的。因此这些数学方法也往往只反映了这些学科的规律，如果死搬硬套去解决别的学科的问题，往往无从下手，甚至导致谬误。为了从根本上解决控制论中的许多问题，他重新研究了数学的基础——集合论，他发现了集合论实质上是扬弃了模糊性而抽象出来的，是把思维过程绝对化，从而达到精确、严格的目的。即一个被讨论的对象 $X$，要么属于某一集合 $A$，记作 $X \in A$，要么不属于该集合，记作 $X \notin A$，二者必居其一，而且二者仅居其一，绝不模棱两可。这种方法完全忽略了 $X$ 对于 $A$ 的隶属程度的差异，但这种差异有时是很重要的。例如命题 $p$：张三是个学生。由于学生这一概念的内涵与外延是明确的，故此命题或取真值“1”，或取假值“0”。但若命题 $Q$：张三的性格稳重。由于“性格稳重”是个模糊概念，其外延是不分明的，怎样判别这一命题的真假呢？

精确性与模糊性的对立，是当今科学发展面临的一个十分突出的矛盾。各门学科迫切要求数学化、定量化，但科学的深入发展意味着研究对象的复杂化，而复杂的东西又往往难以精确化。电子计算机的出现，在一定程度上正在解决着这个矛盾，要求高度的精确，但机器所执行的日益繁杂的任务，往往无法实现高度的精确。例如：命令计算机从监视大厅的摄象镜头中找出

一个长满大胡子的高个子，如果程序在屏幕上提出问题：身长多少以上算大个子，或许你勉强可以回答，但若计算机又问：有多少根胡子以上算大胡子？你将会被这问题弄得啼笑皆非。

决不能将“模糊”两字看成消极的贬义词。过分的精确反而模糊，适当地模糊反而精确。人脑在计算速度、记忆能力等方面远逊于电脑。可电脑对事物的识别远不如人脑，其主要原因是电脑对模糊事物的识别和判决远不如人脑，上述的高个子大胡子，对于人脑来说是远不是什么困难的问题。这二个模糊特征是人所早已掌握好了的，只要把大厅中的人群按对此种特征的隶属程度作比较，即可迅速找到此人。

模糊数学诞生至今仅二十几年的发展历程，它在模式识别这一邻域中的应用历史更为短暂，还远未成熟。本章中尽可能涉及模式识别的本题，并将重点放在隶属度函数的建立上。因为针对某一模式的识别，其难点也正在于此。尽管如此，对于某一特定的模式识别课题，仍没有似乎也不可能提供一种按步就班的通用解法，但我们认为更重要的是解决问题的思路，有了这一基础，就具备了解决具体问题的前提。

## 二、模糊子集

### 1. 模糊子集的定义

1965 年 Zadeh 提出了如下模糊子集的定义：

定义 9.3.1　给定论域 $U$ 上的一个模糊子集 $A$，是指：对于任意 $u\in U$，都确定了一个数 $\mu_A(u)$，称 $\mu_A(u)$ 为 $u$ 对 $A$ 的隶属度，且 $\mu_A(u)\in[0,1]$。

映射：$\mu_A: U\to[0,1]$　　(9.3.1)

$u\to\mu_A(u)$，叫做 $A$ 的隶属函数。

模糊子集完全由其隶属函数所刻划。

当 $\mu_A$ 的值域$=\{0,1\}$时，$\mu_A$ 蜕化为一个普通子集的特征函数，$A$ 便蜕化成一个普通子集。普通子集是模糊子集的特殊形态。

若把论域 $U$ 上全部模糊子集所组成的集合记作 $F(U)$，则有 $F(U)\supseteq P(U)$。

其中 $P(U)$是 $U$ 的幂集。

当 $A\subseteq F(U)-P(U)$时，$A$ 叫真模糊子集。此时，至少存在一元素 $u_0$，使

$$\mu_A(u_0)\notin\{0,1\}。$$

例 9.3.1　如图 9-17 所示 $U=\{a,b,c,d\}$，对 $U$ 的每一元素指定一个它对“圆形”的隶属度，设为

$$\mu_A(a)=1 \qquad \mu_A(b)=0.9$$
$$\mu_A(c)=0.5 \qquad \mu_A(d)=0.2$$

这样便分别表征了它们对于“圆形”的隶属程度。

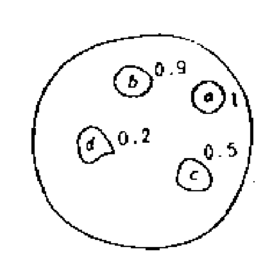


图 9-17　$U$ 上的“圆形”模糊集

当论域 $U$ 是有限集时，可用向量来表示模糊子集 $A$，对于上例可写成 $A=(1,0.9,0.5,0.2)$。

也可以采用 Zadeh 的表示法：

$$A=1/a+0.9/b+0.5/c+0.2/d \qquad (9.3.2)$$

上式右端并不是分式求和，而是一种标记法。分母位置放的是论域 $U$ 中的元素，分子位置放相应元素的隶属度。当某一元素的隶度为 0 时，这一项可以不记入。

也可以采用另一种标记法：

$A=\{(1,a),(0.9,b),(0.5,c),(0.2,d)\}$

例 9.3.2 以年龄作论域，取 $U=[0,200]$，Zadeh 给出了“年老”与“年轻”这两个模糊集 $O,Y$ 的隶属函数如下：

$$\mu_O(u)=\begin{cases}0 & \text{当 } 0\leqslant u<50\\ \left[1+\left(\frac{u-50}{5}\right)^{-2}\right]^{-1} & \text{当 } 50\leqslant u\leqslant 200\end{cases}$$

$$\mu_Y(u)=\begin{cases}1 & \text{当 } 0\leqslant u<25\\ \left[1+\left(\frac{u-2.5}{5}\right)^{-2}\right]^{-1} & \text{当 } 25\leqslant u\leqslant 200\end{cases} \tag{9.3.3}$$

在这一例子中，$U$ 是一个连续的实数区间，$U$ 的模糊子集便可用普通的实函数来表示，当 $U$ 从有限论域推广到一般时，可采用 Zadeh 如下的标记法：

$$A=\int_U(\mu_A(u)/u) \tag{9.3.4}$$

这与式(9.3.2)的表示法基本相同，这里的积分号和式(9.3.2)一样也不是求和号，而是表示各个元素与隶属度对应关系的一个总括。

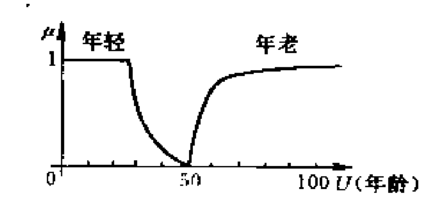



图 9-18 “年轻”“年老”的隶属函数曲线

2. *隶属函数的确定*

隶属度是模糊集合赖以建立的基石，要确定恰当的隶属函数并不容易，迄今仍无一个“放之四海而皆准”的法则可遵循。这需要对被描述的概念有足够的了解，一定的数学技巧，而且还包括心理测量的进行与结果的运用。正如某一事件的发生与否有一定的不确定性(随机性)一样，某一对象是否符合某一概念也有一定的不确定性(我们称之为模糊性)。

随机性是因果律的一种破缺。在那里事件本身具有明确的含义，只是由于条件不完全，使得在条件与事件间不能出现决定性的因果关系。概率论的运用，得以从随机性中把握广义的因果律——概率规律。

模糊性则是排中律的一种破缺，在这里由于概念本身没有明确的外延，故而某一对象是否符合这一概念的划分，就有不确定性，模糊数学正是从这一不确性(模糊性)中确立广义的排中律——隶属规律。

诚然，隶属度的具体确定，往往包含着人脑的加工，包含着某种心理过程。但心理过程也是物质性的，心理物理学的大量实验已经表明：人由各种感觉获得的心理量与外界刺激的物理量间，保持着相当严格的关系，对心理测量结果的运用与修正，导致了隶属度的正确建立。

(1)模糊统计法：

在某些场合下，隶属度可用模糊统计的方法来确定。读者对于概率统计当然是熟悉的，建议在阅读下面介绍的模糊统计时，将它与概率统计的异同作一比较，以加深理解。

模糊统计试验，有四个要素：

①论域 $U$，例如人的集合；

②$U$ 中的一个元素，例如李平；

③$U$ 中一个边界可变的普通集合 $A^*$，例如“高个子”。$A^*$ 联系于一个模糊集 $A$ 及相应的模糊概念 $\alpha$；

④条件 $s$，它联系着按概念 $\alpha$ 所进行的划分过程的全部主客观因素，它制约着 $A^*$ 边界的改

变。例如不同试验者对“高个子”的理解。

模糊性产生的根本原因是：$s$ 对按概念 $\alpha$ 所作的划分引起 $A^*$ 的变异，它可能覆盖了 $u_0$，也可能不覆盖 $u_0$，这就导致 $u_0$ 对 $A^*$ 的隶属关系不确定。例如有的试验者认为李平是“高个子”，但有的试验者认为他不是。

模糊统计试验的基本要求是在每一次试验下，要对 $u_0$ 是否属于 $A^*$ 作一个确切的判断，作 $n$ 次试验（即让 $n$ 位试验者对李平是否属于“高个子”作判断），就可算出 $u_0$ 对 $A$ 的隶属频率：

$$u_0 \text{对} A \text{的隶属频率} \triangleq \frac{“u_0 \in A^*”\text{的次数}}{n} \qquad (9.3.5)$$

式中≙这一记号表示“记作”，“定义为”。

许多实验证明，随着 $n$ 的增大，隶属频率呈现稳定性，被称为隶属频率稳定性，频率稳定所在的数值叫 $u_0$ 对 $A$ 的隶属度。即有

$$\mu_A(u_0) = \lim_{n \to \infty} \frac{“u_0 \in A^*”\text{的次数}}{n} \qquad (9.3.6)$$

上例中若在 100 位试验者中有 90 位认为李平是“高个子”，则可认为 $\mu_{\text{高个子}}(\text{李平}) = 0.9$，即李平对于“高个子”的隶属度为 0.9。

用模糊统计这一方法，对“青年人”这一概念是适宜的年龄作抽样试验，得到很好的结果。见图 9-19。

(2)二元对比排序法：

人们习惯于从两事物的对比中，作出它们对某一概念符合程度的判断，例如说茶花比月季花好看。可惜这种判断往往不满足数学上对“序”的要求，往往不具有传递性，而出现循环的现象。例如今天认为甲花比乙花好看，乙花比丙花好看，但当明天单独将甲、丙两种花放在一起时，很可能认为丙花比甲花好看。问题恰恰是因为“好看”这一模糊概念引起的，影响“好看”与否的因素太多。

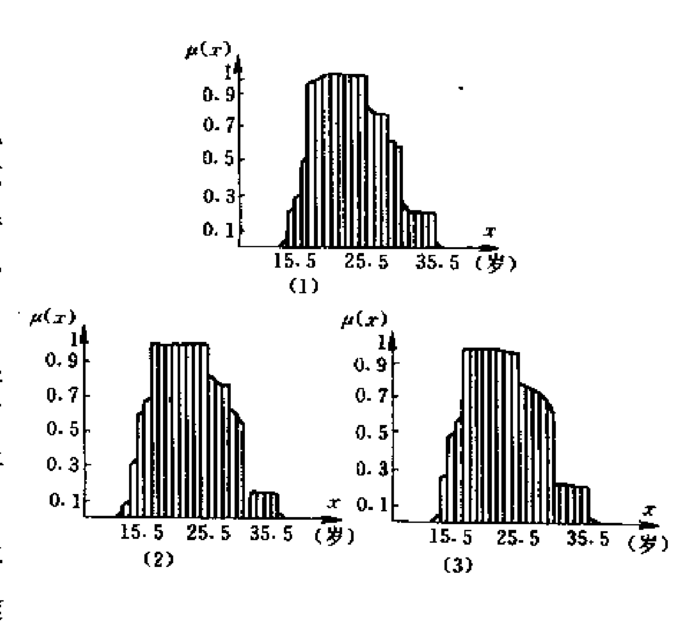


图 9-19 “青年人”的年龄隶属函数曲线

但二元对比毕竟是区别事物的重要方法，下面介绍据此建立隶属函数的途径。

①择优比较法　此方法可以下例说明：

例 9.3.3　求茶花、月季、牡丹、梅花、荷花对“好看的花”的隶属度。

选 10 名试验者，令之逐次对二种花作对比，并赋予优胜者 1 分，失败者 0 分，故每一试验者需作 $C_5^2=10$ 次对比，设某一试验者的对比结果如表 9-5 所示，则累计 10 位试验者的结果便可求出该种花的总得分，从而得到该种花对“好看的花”的隶属度。见表 9-6。

表 9-5　一位试验者的二元对比结果

| 优胜＼失败 | 茶花 | 月季 | 牡丹 | 梅花 | 荷花 | 得分 |
| --- | --- | --- | --- | --- | --- | --- |
| 茶花 | | 1 | 0 | 1 | 0 | 2 |
| 月季 | 0 | | 0 | 1 | 0 | 1 |
| 牡丹 | 1 | 1 | | 1 | 0 | 3 |
| 梅花 | 0 | 0 | 0 | | 0 | 0 |
| 荷花 | 1 | 1 | 1 | 1 | | 4 |

表 9-6　5 种花对“好看的花”的隶属度

| 名称 | 总得分 | 隶属度 |
| --- | --- | --- |
| 茶花 | 23 | 0.23 |
| 月季 | 18 | 0.18 |
| 牡丹 | 20 | 0.20 |
| 梅花 | 15 | 0.15 |
| 荷花 | 24 | 0.24 |

②相对比较法　设欲对论域 $U$ 中的元素 $x$、$y$、…，按某种特性排序，先在两两元素的对比中建立比较值，例如取元素 $x$、$y$ 作比较，则得到比较值 $f_y(x)$、$f_x(y)$，其意义为：若 $x$ 与 $y$ 对比，$x$ 具有某特性的程度具有值：$f_y(x)$；则 $y$ 与 $x$ 对比，$y$ 具有该特性的程度具有值：$f_x(y)$。

例如：长子与次子相比较，若把长子具有与父亲的相似程度 $f_y(x)$ 定为 0.8，则可把次子具有与父亲相似的程度 $f_x(y)=0.5$，这里 0.8，0.5 均不是他们与父亲相似程度的绝对度量，而是说若长子像父亲八分的话，则次子像父亲仅五分。当然若把长子像父亲定为九分，则次子像父亲也可能有六分。

在这种两元素相对比较取值基础上，可通过一定算法得到总体排序。

例 9.3.4　某人有长、次、幼三子（分别记作 $x$，$y$，$z$），且有与父亲的相对相似程度：

$$(f_y(x), f_x(y)) = (0.8, 0.5)$$
$$(f_z(y), f_y(z)) = (0.4, 0.7)$$
$$(f_x(z), f_z(x)) = (0.3, 0.5)$$
$$f_x(x) = f_y(y) = f_z(z) = 1$$

求它们的总体排序。

为此令

$$f(x/y) = \begin{cases} f_y(x)/f_x(y), & \text{当 } f_y(x) \leqslant f_x(y) \\ 1 & \text{当 } f_y(x) > f_x(y) \end{cases}$$

则可得相及矩阵

$$\begin{bmatrix} 1 & 1 & 1 \\ 5/8 & 1 & 4/7 \\ 3/5 & 1 & 1 \end{bmatrix}$$

在相及矩阵每行中取最小值，得向量

$$\begin{bmatrix}1\\4/7\\3/5\end{bmatrix}$$

则其排序为:长、幼、次(按与父亲的相似程度递减排序)。

③推理法　在某些应用场合下,隶属度函数可作为一种推理的产物出现。例如评价一封闭曲线的圆度,可根据该封闭曲线的周长相对于内切圆周长的裕度来评价。又如根据某一线段与水平线的交角来分辨横竖撇捺等等。

设计者在不同的应用场合,可根据不同的数学物理知识,设计出隶属度函数,然后在实践中检验调整之。但在很多应用课题上,很难用推理法获得隶属度函数,对此必需有清醒的估计。下面提供一些实例,以供设计隶属函数时借鉴。

例 9.3.5　三角形的隶属函数。

在染色体自动识别或白血球分类等课题中,常常把问题归结为几何图形的识别,而几何图形总可近似为若干凸多边形,而凸多边形总可近似为若干三角形的合成。故有必要判断一三角形属于等腰三角形($I$),直角三角形($R$),等腰直角三角形($IR$),正三角形($E$),或不是上述三角形($T$)。

设三角形三内角分别为 $A$、$B$、$C$,且 $A\geqslant B\geqslant C\geqslant 0$,则可规定

$$\mu_I(A,B,C)=1-\frac{1}{60}\min(A-B,B-C) \tag{9.3.7}$$

因为当 $A=B$,或 $B=C$ 时该三角形为等腰三角形,$\mu_I=1$;而当 $A=120,B=60,C=0$(最不等腰),$\mu_I=0$。

$$\mu_R(A,B,C)=1-\frac{1}{90}(A-90) \tag{9.3.8}$$

因为当 $A=90$ 时,$\mu_R=1$;而当 $A=180,B=C=0$ 时,$\mu_R=0$。

$$\mu_E(A,B,C)=1-\frac{1}{180}(A-C) \tag{9.3.9}$$

因为当 $A=B=C,\mu_E=1$;而当 $A=180,B=C=0$ 时,$\mu_E=0$。

因　　$IR=I\cap R$

故 $$\mu_{IR}(A,B,C)=\min\left[1-\frac{1}{60}\min(A-B,B-C),1-\frac{1}{90}|A-90|\right] \tag{9.3.10}$$

因为 $T=\bar{I}\cap\bar{E}\cap\bar{R}$　故

$$\mu_T(A,B,C)=\min[1-\mu_I(A,B,C),1-\mu_E(A,B,C),1-\mu_R(A,B,C)] \tag{9.3.11}$$

例 9.3.6　笔划类型的隶属函数

汉字中横竖撇捺等线段的区分是根据它们与水平线的交角确定的。

设 $A$ 为一线段,$H,V,S,BS$ 为横竖撇捺四模糊集,则

$$\mu_H(A)=1-\min\left(\frac{|\theta|}{45},1\right) \tag{9.3.12}$$

$$\mu_V(A)=1-\min\left(\frac{|90-\theta|}{45},1\right) \tag{9.3.13}$$

$$\mu_S(A)=1-\min\left(\frac{|45-\theta|}{45},1\right) \tag{9.3.14}$$

$$\mu_{BS}(A)=1-\min\left(\frac{|135-\theta|}{45},1\right) \tag{9.3.15}$$

例 9.3.7　手写体字符"$U$","$V$"之区别。

手写体大写字符“$U$”、“$V$”常被划分到同一类中，它们的进一步区别可用隶属度函数实现。考虑到手写字符“$V$”的两边总是比“$U$”的两边平直，故可用它们的图形所包含的面积与三角形面积 $S\left(\frac{1}{2}\times\text{底边长}\times\text{高}\right)$ 作比较，接近三角形面积者为“$V$”，否则为“$U$”。

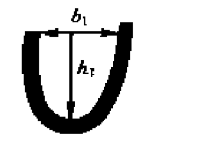





图 9-20　手写字符“$U$”与“$V$”

三角形底边长 $b$ 与高 $h$ 如图 9-20 所示，字符所包含的内面积 $S'$ 定义为上述底边线与字符内侧所包含的面积。据此设计的隶属度函数 $\mu_V$ 定义如下：

$$\mu_V = 1 - \left|\frac{S'}{\frac{1}{2}bh}\right| \tag{9.3.16}$$

统计表明 $\mu_V > 0.8$ 时，应判决为“$V$”，否则判决于“$U$”。

## 三、模糊关系

本节中将介绍模糊关系的定义、建立及其运算。当论域有限时，模糊关系可用矩阵表示，这就导致讨论模糊矩阵的理论及其在模式分类中的应用。

### 1. 模糊关系的性质及其建立

（1）模糊关系的性质　设 $U$、$V$ 是两论域，记

$$U\times V=\{(x,y)\mid x\in U, y\in V\}$$

为 $U$ 与 $V$ 的笛卡尔乘积集。

笛卡尔乘积集是两集合元素间的无约束搭配。若给搭配以约束，便体现了一种特殊关系，接受此种约束的元素对构成笛卡尔乘积集的一个子集，该子集便表现了一种关系。因此在普通集合论中，所谓 $U$ 到 $V$ 的一个关系，乃是被定义为 $U\times V$ 的一个子集 $R$。

$$R\in F(U\times V) \quad \text{记作} \quad U\xrightarrow{R}V。$$

定义 9.3.2　称 $U\times V$ 的一个模糊子集 $R$ 为从 $U$ 到 $V$ 的一个模糊关系。记作 $U\xrightarrow{R}V$。

模糊关系 $R$ 的隶属函数

$$\mu_R: U\times V\rightarrow[0,1] \tag{9.3.17}$$

$\mu_R(x_0,y_0)$ 叫做 $(x_0,y_0)$ 具有关系 $R$ 的程度。

当论域 $U$、$V$ 都是有限论域，此时模糊关系 $R$ 可以用矩阵 $R$ 表示，即

$$R=(r_{ij}) \tag{9.3.18}$$

其中 $r_{ij}=\mu_R(x_i,y_j)$

显然有

$$0\leq r_{ij}\leq 1 \quad (1\leq i,j\leq n) \tag{9.3.19}$$

满足式(9.3.18)、式(9.3.19)的矩阵，称作模糊矩阵，特别地，当

表 9-7　苹果、乒乓等七种物品的相似程度

| 相似度 | 苹果 | 乒乓球 | 书 | 足球 | 桃子 | 气球 | 四棱锥 |
|---|---|---|---|---|---|---|---|
| 苹　果 | 1 | 0.7 | 0 | 0.7 | 0.5 | 0.6 | 0 |
| 乒乓球 | 0.7 | 1 | 0 | 0.9 | 0.4 | 0.5 | 0 |
| 书 | 0 | 0 | 1 | 0 | 0 | 0 | 0.1 |
| 足　球 | 0.7 | 0.9 | 0 | 1 | 0.4 | 0.5 | 0 |

续表

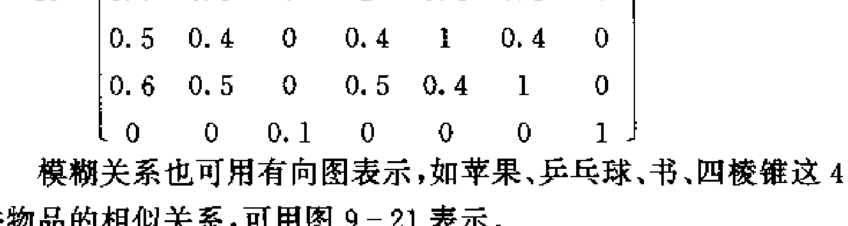
| 相似度 | 苹果 | 乒乓球 | 书 | 足球 | 桃子 | 气球 | 四棱锥 |
|---|---|---|---|---|---|---|---|
| 桃　子 | 0.5 | 0.4 | 0 | 0.4 | 1 | 0.4 | 0 |
| 气　球 | 0.6 | 0.5 | 0 | 0.5 | 0.4 | 1 | 0 |
| 四棱锥 | 0 | 0 | 0.1 | 0 | 0 | 0 | 1 |

$$r_{ij}\in\{0,1\}\qquad(1\leqslant i,j\leqslant n)\tag{9.3.20}$$

则矩阵 $R$ 退化为布尔矩阵。布尔矩阵可以表达一种普通关系。

例 9.3.8　用模糊关系表示苹果、乒乓球、书、足球、桃子、气球、四棱锥的相似关系，设用专家评分的办法给出它们的相似程度如下表，显然，其相似矩阵为

$$R=\begin{bmatrix}1 & 0.7 & 0 & 0.7 & 0.5 & 0.6 & 0\\ 0.7 & 1 & 0 & 0.9 & 0.4 & 0.5 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0.1\\ 0.7 & 0.9 & 0 & 1 & 0.4 & 0.5 & 0\\ 0.5 & 0.4 & 0 & 0.4 & 1 & 0.4 & 0\\ 0.6 & 0.5 & 0 & 0.5 & 0.4 & 1 & 0\\ 0 & 0 & 0.1 & 0 & 0 & 0 & 1\end{bmatrix}$$



模糊关系也可用有向图表示，如苹果、乒乓球、书、四棱锥这 4 件物品的相似关系，可用图 9-21 表示。

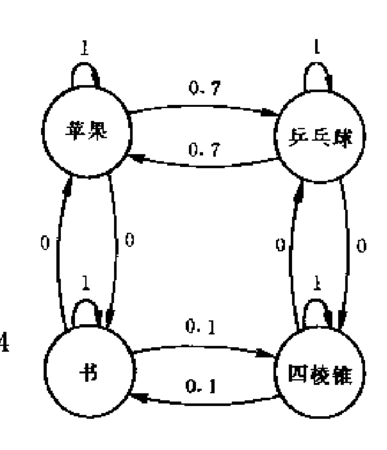



图 9-21　模糊关系图

(2)模糊关系的建立

读者从后面的介绍将会看到模糊关系在聚类分析中的应用，但对样品分类的效果怎样？关键是选择合理的统计指标，即被选中的指标应有明显的实际意义，有较强的分辨力和代表性。在统计指标选定后即可按下述步骤建立模糊关系，进而运用后面介绍的内容进行分类。

第一步：把各代表点的统计指标的数据标准化，以便分析和比较。这一步也称正规化。

为把标准化数据压缩[0,1]闭区间，可用极值标准化公式

$$x=\frac{x'-x'_{\min}}{x'_{\max}-x'_{\min}}\tag{9.3.21}$$

当 $x'=x_{\max}$时，$x=1$；当 $x'=x'_{\min}$时，$x=0$，否则取[0,1]之间。

第二步：算出被分类对象间具有此种关系的程度 $r_{ij}$(最通常是 $i$ 与 $j$ 的相似程度)，其中 $i$、$j=1、2、\cdots、n$，$n$ 为对象个数。从而确定论域 $U$ 上的模糊关系 $R$

$$R=\begin{bmatrix}r_{11} & r_{12} & \cdots & r_{1n}\\ r_{21} & r_{22} & \cdots & r_{2n}\\ \vdots & & & \\ r_{n1} & r_{n2} & \cdots & r_{nn}\end{bmatrix}\tag{9.3.22}$$

计算 $r_{ij}$的常用方法有：

①欧氏距离法

$$r_{ij}=\sqrt{\frac{1}{m}\sum_{k=1}^{m}(x_{ik}-x_{jk})^2}\tag{9.3.23}$$

式中 $x_{ik}$为第 $i$ 个对象第 $k$ 个因子的值，$x_{jk}$为第 $j$ 个对象第 $k$ 个因子的值。

②数量积

$$r_{ij}=\begin{cases}1 & 当\ i=j\\ \sum_{k=1}^{m}\frac{x_{ik}\cdot x_{jk}}{M} & 当\ i\neq j\end{cases} \tag{9.3.24}$$

其中 $M$ 为一适当选择之正数，满足

$$M\geqslant \max_{ij}\left(\sum_{k=1}^{m}x_{ik}\cdot x_{jk}\right) \tag{9.3.25}$$

③相关系数

$$r_{ij}=\frac{\sum_{k=1}^{m}|x_{ik}-\overline{x}_i||x_{jk}-\overline{x}_j|}{\sqrt{\sum_{k=1}^{m}(x_{ik}-\overline{x}_i)^2}\cdot\sqrt{\sum_{k=1}^{m}(x_{jk}-\overline{x}_j)^2}} \tag{9.3.26}$$

其中

$$\overline{x}_i=\frac{1}{m}\sum_{k=1}^{m}x_{ik},\qquad \overline{x}_j=\frac{1}{m}\sum_{k=1}^{m}x_{jk}$$

④指数相似系数

$$r_{ij}=\frac{1}{m}\sum_{k=1}^{m}\exp\left(-\frac{3}{4}\frac{(x_{ik}-x_{jk})^2}{S_k^2}\right) \tag{9.3.27}$$

其中 $S_k$ 为适当选择之正数。

⑤非参数法

令 $x'_{ik}=x_{ik}-\overline{x}_i$

$n^{+}=\{x'_{i1}\cdot x'_{j1}\cdot x'_{i2}\cdot x'_{j2},\cdots,x'_{im}\cdot x'_{jm}\}$中大于 0 的个数

$n^{-}=\{x'_{i1}\cdot x'_{j1}\cdot x'_{i2}\cdot x'_{j2},\cdots,x'_{im}\cdot x'_{jm}\}$中小于 0 的个数

$r_{ij}=\frac{|n^{+}-n^{-}|}{n^{+}+n^{-}}$

⑥最大最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\sum_{k=1}^{m}\max(x_{ik},x_{jk})} \tag{9.3.28}$$

⑦算术平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\frac{1}{2}\sum_{k=1}^{m}(x_{ik}+x_{jk})} \tag{9.3.29}$$

⑧几何平均最小法

$$r_{ij}=\frac{\sum_{k=1}^{m}\min(x_{ik},x_{jk})}{\sum_{k=1}^{m}\sqrt{x_{ik}\cdot x_{jk}}} \tag{9.3.30}$$

⑨绝对值指数法

$$r_{ij}=\exp\left(-\sum_{k=1}^{m}|x_{ik}-x_{jk}|\right) \tag{9.3.31}$$

⑩绝对值倒数法

$$r_{ij}=\begin{cases}1 & 当\ i=j\\ \dfrac{M}{\sum\limits_{k=1}^{m}|x_{ik}-x_{jk}|} & 当\ i\neq j\end{cases}\qquad(9.3.32)$$

其中 $M$ 适当选取，使 $0\leqslant r_{ij}\leqslant 1$。

2. 基于模糊等价关系的模式分类

在模糊矩阵的运算中提出了模糊关系的自反性、对称性、传递性，并指出同时满足这三种性质的关系才是模糊等价关系。

例 9.3.9 设有 5 种矿石，按其颜色、比重等性质得出描述其“相似程度”的模糊关系矩阵如下：

$$R=\begin{array}{c}\\ x_1\\ x_2\\ x_3\\ x_4\\ x_5\end{array}\begin{array}{c}\begin{array}{ccccc}x_1 & x_2 & x_3 & x_4 & x_5\end{array}\\ \begin{bmatrix}1 & 0.8 & 0 & 0.1 & 0.2\\ 0.8 & 1 & 0.4 & 0 & 0.9\\ 0 & 0.4 & 1 & 0 & 0\\ 0.1 & 0 & 0 & 1 & 0.5\\ 0.2 & 0.9 & 0 & 0.5 & 1\end{bmatrix}\end{array}$$

本矩阵的自反性与对称性是明显的，下面的计算也可证明它不具有传递性，即它仅是相似矩阵，先不加改造即用来分类。

若认为彼此“相似程度”大于 0.8 的为一类，则 $x_1$、$x_2$ 为一类，$x_2$、$x_5$ 为一类，但 $x_1$、$x_5$ 的“相似程度”仅 0.2，故 $x_1$、$x_5$ 不属一类，这样就得到矛盾，说明模糊相似矩阵不能直接用来分类。为了得到模糊等价关系，可用 $R$ 自乘得 $R^2$。即 $R\circ R=R^2$，$R^2\circ R^2=R^4$，…直到 $R^{2k}=R^k$。至此，$R^k$ 便是一模糊等价关系。此方法是由“传递闭包”而来的，此处不作证明。

在本例中 $R^2\neq R$，故 $R$ 不是模糊等价矩阵，因

$$R^2=R\circ R=\begin{bmatrix}1 & 0.8 & 0.4 & 0.2 & 0.8\\ 0.8 & 1 & 0.4 & 0.5 & 0.9\\ 0.4 & 0.4 & 1 & 0 & 0.4\\ 0.2 & 0.5 & 0 & 1 & 0.5\\ 0.8 & 0.9 & 0.4 & 0.5 & 1\end{bmatrix}$$

类似地还可求出 $R^4$ 与 $R^8$，并且它们是相等的，故 $R^4$ 就满足模糊等价关系的性质。

定理 9.3.1 $R\in M_{n\times n}$是等价矩阵，当且仅当任意 $\lambda\in[0,1]$，$R_\lambda$ 都是等价的布尔矩阵。

根据定理 9.3.1 可知：若 $R$ 为模糊等价关系，则对于给定的 $\lambda\in[0,1]$便可得到相应的普通等价关系 $R_\lambda$，这意味着得到了一个 $\lambda$ 水平的分类。

定理 9.3.2 若 $0\leqslant\lambda<\mu\leqslant 1$，则 $R_\mu$ 所分出的每一类必是 $R_\lambda$ 所分出的某一类的子类，或称 $R_\mu$ 的分类法是 $R_\lambda$ 分类法的“加细”。

证：

$$\begin{aligned}r_{ij}^{\mu}=1 &\Leftrightarrow r_{ij}\geqslant\mu\\ &\Rightarrow r_{ij}>\lambda\\ &\Leftrightarrow r_{ij}^{\lambda}=1\end{aligned}$$

亦即 $r_{ij}^{\mu}=1\Rightarrow r_{ij}^{\lambda}=1\qquad(\lambda<\mu)$

这说明，若 $i,j$ 按 $R_\mu$ 能被归为一类，则按 $R_\lambda$ 必被归为一类。

当 $\lambda$ 自 1 逐渐降为 0，则其决定的分类逐渐变粗，逐步归并，形成一动态的聚类图。

例 9.3.10 设论域

$$U=\{x_1,x_2,x_3,x_4,x_5\}$$

给定模糊关系矩阵

$$R=\begin{bmatrix}1 & 0.48 & 0.62 & 0.41 & 0.47\\ 0.48 & 1 & 0.48 & 0.41 & 0.47\\ 0.62 & 0.48 & 1 & 0.41 & 0.47\\ 0.41 & 0.41 & 0.41 & 1 & 0.41\\ 0.47 & 0.47 & 0.47 & 0.41 & 1\end{bmatrix}$$

其自反性与对称性是显然的，经验证可知 $R\circ R\subseteq R$，故 $R$ 为一模糊等价关系。

现根据不同的 $\lambda$ 水平分类：

(1) 当 $0.62<\lambda\leqq 1$ 时

$$R_\lambda=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

此时共分为 5 类：$\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\}$，即每个元素为一类，这是“最细”的分类。

(2)当 $0.48<\lambda\leqslant 0.62$ 时

$$R_\lambda=\begin{bmatrix}1&0&1&0&0\\0&1&0&0&0\\1&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

此时共分四类：$\{x_1,x_3\},\{x_2\},\{x_4\},\{x_5\}$。

(3)当 $0.47<\lambda\leqslant 0.48$ 时

$$R_\lambda=\begin{bmatrix}1&1&1&0&0\\1&1&1&0&0\\1&1&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

此时共分三类：$\{x_1,x_2,x_3\},\{x_4\},\{x_5\}$。

(4)当 $0.41<\lambda\leqslant 0.47$ 时

$$R_\lambda=\begin{bmatrix}1&1&1&0&1\\1&1&1&0&1\\1&1&1&0&1\\0&0&0&1&0\\1&1&1&0&1\end{bmatrix}$$

此时共分两类：$\{x_1,x_2,x_3,x_5\},\{x_4\}$。

(5)当 $0\leqslant\lambda\leqslant0.41$ 时，$R_\lambda$ 的元素全为 1，故 5 元素合为一类，即是“最粗”的分类。

综合上述结果，可画出动态聚类图(图 9－22)，这也是一个基于模糊等价关系完成聚类分析的实例。

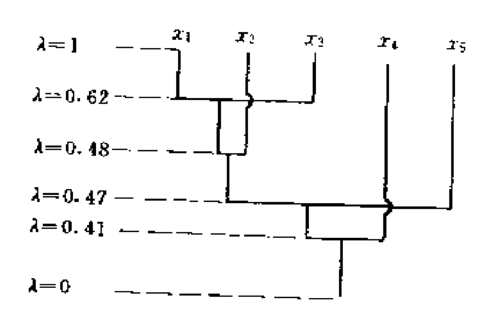



图 9－22　按 λ 的不同水平进行聚类

3. 基于模糊相似关系的分类

例 9.3.9　以一反例表明，对于仅具有自反性与对称性的模糊相似关系，需改造成为模糊等价关系，才能进行正确分类。但由于多次合成操作，消耗机时很多，特别当元素个数很多时，这一问题变得更严重，为此众多学者纷纷寻找由模糊相似矩阵直接进行聚类的方法，如最大树法编网法。现仅将最大树法介绍于下。

最大树法：先画出被分类的元素集。从矩阵 $R$ 中按 $r_{ij}$ 从大到小的顺序依次连边，标上权重。若在某一步会出现回路，便不画那一步，直到所有元素连通为止，这样就得到一棵“最大树”(可以不唯一)。取定 $\lambda$，砍去权重低于 $\lambda$ 的边，便可将元素分类，互相连通的元素归为同类。下面以日本学者 Tamura 的例子来说明之。

例 9.3.11　设有三个家庭，每家 4～7 人，选每个人的一张照片，共 16 张混放在一起，请中学生对照片两两比较，按相似程度聚类，希望能把三个家庭区分开。

16 张照片的相似矩阵见表 9－8，现以此例来构造最大树：设先选顶点 $i=1$，依次连“13”标 $r_{ij}=0.8$ 于其边侧，再连“16”，$r_{ij}=0.6$，由“16”连接“6”，$r_{ij}=0.8$，…，依此下去得到一棵连通 16 个顶点的最大树，如图 9－23 所示。

表 9－8　16 张照片的相似矩阵

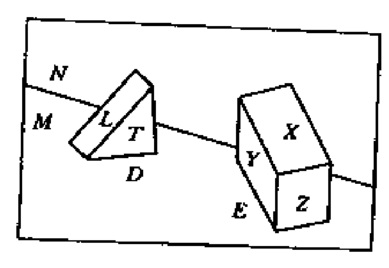
| $r_{ij}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | | | | | | | | | |
| 2 | 0 | 1 | | | | | | | | | | | | | | |
| 3 | 0 | 0 | 1 | | | | | | | | | | | | | |
| 4 | 0 | 0 | 0.4 | 1 | | | | | | | | | | | | |
| 5 | 0 | 0.8 | 0 | 0 | 1 | | | | | | | | | | | |
| 6 | 0.5 | 0 | 0.2 | 0.2 | 0 | 1 | | | | | | | | | | |
| 7 | 0 | 0.8 | 0 | 0 | 0.4 | 0 | 1 | | | | | | | | | |
| 8 | 0.4 | 0.2 | 0.2 | 0.5 | 0 | 0.8 | 0 | 1 | | | | | | | | |
| 9 | 0 | 0.4 | 0 | 0.8 | 0.4 | 0.2 | 0.4 | 0 | 1 | | | | | | | |
| 10 | 0 | 0 | 0.2 | 0.2 | 0 | 0 | 0.2 | 0 | 0.2 | 1 | | | | | | |
| 11 | 0 | 0.5 | 0.2 | 0.2 | 0 | 0 | 0.8 | 0 | 0.4 | 0.2 | 1 | | | | | |
| 12 | 0 | 0 | 0.2 | 0.8 | 0 | 0 | 0 | 0 | 0.4 | 0.8 | 0 | 1 | | | | |
| 13 | 0.8 | 0 | 0.2 | 0.4 | 0 | 0.4 | 0 | 0.4 | 0 | 0 | 0 | 0 | 1 | | | |
| 14 | 0 | 0.8 | 0 | 0.2 | 0.4 | 0 | 0.8 | 0 | 0.2 | 0.2 | 0.6 | 0 | 0 | 1 | | |
| 15 | 0 | 0 | 0.4 | 0.8 | 0 | 0.2 | 0 | 0 | 0.2 | 0 | 0 | 0.2 | 0.2 | 0 | 1 | |
| 16 | 0.6 | 0 | 0 | 0.2 | 0.2 | 0.8 | 0 | 0.4 | 0 | 0 | 0 | 0 | 0.4 | 0.2 | 0.4 | 1 |

然后对最大树取 $\lambda$ 截集，即去掉那些 $r_{ij}<\lambda$ 的边，这样就可将它截成互不连通的几棵子树。

现取 $\lambda=0.5$，可得三棵子树，其顶点集为

$$V_1=\{13,1,16,6,8,4,9,15,12,10\}$$

$$V_2=\{3\}$$



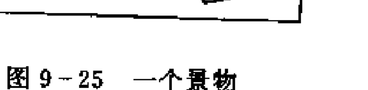


图 9－25　一个景物

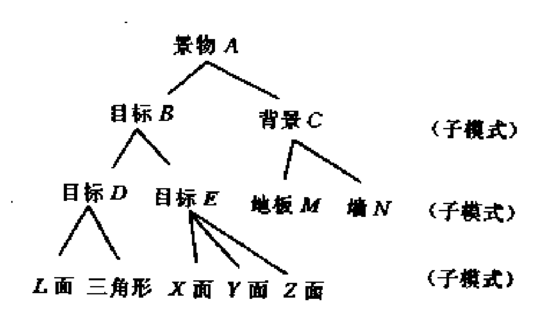



图 9－26　图 9－25 的层次结构描述

是否遵守指定的语法。与此同时，语法分析还产生出该句子的结构描述，通常是树状的结构描述。

结构模式识别的方法，提供了用小而简单的基元与语法规则来描述大而复杂的模式的能力。进一步分析表明，这种方法还有一个引人入胜的特点，这就是利用了语法的递归性。因为同一语法规则，可以递归地应用任意多次，所以它就能以十分紧凑的形式，表示一个甚大的句子集合。当然这样一种方法的实用程度和基元的辨识能力有关，还和用合成操作表达基元间相互关系的能力有关。

通常可用“与”、“或”这样的逻辑操作算符来对子模式作合成操作。例如，我们仅选择“链接”为描述模式的唯一关系，且选取图 9－27(a)中所示的基元，则图 9－27(b)所示的方框可表示为串：*aaabbccccdd*。

进而，若用“＋”表示“从头到尾的链接”这样一种操作，则图 9－27(b)的矩形可表示为 $a+a+a+b+b+c+c+c+d+d$。仍采用图 9－27(a)中的基元来表示图 9－28(a)所示的模式，则如图 9－28(b)所示。可以看出，在这种描述方法下，基元与基元之间的关系均被表示为串中的符号。

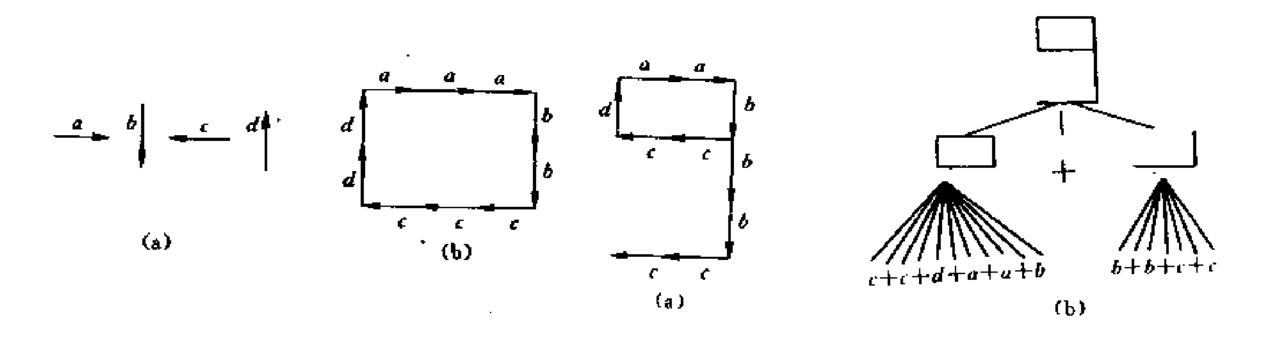



图 9－27　一个矩形及其模式基元

图 9－28　模式及其结构描述

一种模式的结构信息，也可以用“关系图”来表示。由于在关系图与矩阵间存在一一对应关系，一个关系图就无疑可表示为一个“关系矩阵”。在模式描述中采用关系图，就能扩充允许的关系的种类，以包含那些从模式中很容易确定的关系。(注意：①链接是一维语言唯一的基本操作；②一个图可以包含闭环，而树不允许有闭环。)用这种扩展了的形式，我们就能拥有比树状结构更强的描述手段。然而采用树状结构，却提供了一种直接的方法，使得正规的语言理论适应紧凑地描述问题的要求，并适于分析含有结构内容的模式。

## 二、结构模式识别系统

一个结构模式识别系统可认为由三个主要部分组成，它们分别是：预处理、模式描述、语法分析。该系统的简单框图如图 9－29 所示。

每一经过预处理的模式，在模式描述阶段被表示成类语句结构（例如一个串、一个图）。这一阶段的处理包括两部分：①模式分割；②基元抽取。为了将一个模式用若干子模式来表示，我们应该对该模式予以分割，与此同时，抽取基元和基元间的关系。换言之，每一经过预处理的模式，均按照预先制定的语法或合成操作规则，分割为子模式。进而，每一子模式分割为一组基元。当然，若待识模式较简单，则也可以直接分割为基元。例如，借助于链接操作，每一模式可用一串链接起来的基元表示。在语法分析阶段，系统应对已完成描述的模式作语法检查，以判定它是按何种语法结合成的，从而完成待识模式的分类。

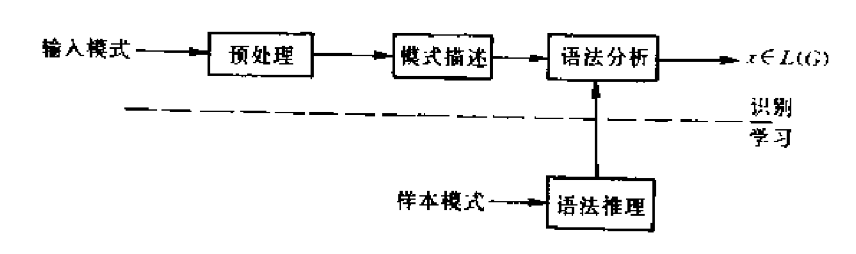


图 9－29 一个结构模式识别系统的框图

一般说来，识别的最简单的方式可能是"样板匹配"。用表示一种输入模式的基元串与各种模型（也是用基元串表示的）相比较，按照选定的匹配准则，输入模式被划入匹配"最好"的那一类。在这种"样板匹配"方式下，层次结构信息基本上不予考虑。上述"样板匹配"是一种方法，另一种识别方法是研究全部层次结构信息。此外还有若干介于这两种方法之间的研究方法。例如，设计一系列测试，以测定某些子模式（或某些基元）是否存在，或测定某些子模式的特定组合（或某些基元的特定组合）是否存在，测定的结果，用作分类判决。注意，这里所说的一次测试，可以是一次样板匹配，也可以是对表达一个子模式的一棵子树的分析。要知道：识别方法的选取，通常取决于待识别的模式，若识别要求完整的模式描述，就要分析全部层次结构信息，反之就可用较简单的方法提高识别过程的效率，避免作完整的层次结构分析。

一类模式的结构信息，要有一个文法来描述（实际上就是描述该类的结构，以与其它类相区别），这就需要文法推理，它能从给它的训练模式集中，归纳出一个文法，这类似于统计模式识别中用样本来训练判决函数。从图 9－29 可以看出，一类样品被送入语法推理机构，形成了该类的结构描述，作为语法分析机构入口之一的"样板"（如果按前面所说用样板匹配的话），或是层次结构分析的"文法"（如果用前述的层次结构分析），其实，更广义的学习，还应包括基元的位置选定，这也是该推理机构的一种能力。

## 三、模式基元的选择与抽取

### 1. 模式基元的选择

在模式描述中，为构成一个句子模型，首要的一步是要确定一个模式基元集，但这受制于数据本身的特征、用途、技术可行性等因素。没有一成不变的办法来解决基元选择问题，这里只能提出一些方向性的建议：

①所选取的基元应是精简的，以易于语法描述与分析。

②能用非语言学的方法方便地抽取，因为它们被认为是最基本的模式单元。

例如，就语音识别而言，通常认为对链接式关系来说，音素就是应被选取的模式基元。类似地，对文字识别而言，当然是选取笔划了。但对于一般图片来说，就没有类似于音素或笔划这样的通用的基元了。有时为了能对模式进行充分的描述，所选的基元可能只对那些专门用途才是重要的，例如若尺寸(或形状、位置)在待识的问题中是重要的，这时就应选取与尺寸(或形状、位置)有关的基元。类似这样的要求通常会导致要选用一些有语义的特性的基元。

下面这一简单的例子可说明这一点，对于同一数据，由于问题的描述不同，导致选用的基元也不同。

例 9.4.2　若问题是从非矩形中挑出矩形，则可选用下述基元：

$a^1$——0°水平线段

$b^1$——90°垂直线段

$c^1$——180°水平线段

$d^1$——270°垂直线段

这样，不论多大的矩形，均被表示为一种语句，即 $a^1b^1c^1d^1$。

但若还要求确定矩形的尺寸，则上述基元就不够了，这时应选单位长度的线段为基元，这时不同尺寸的矩形就会表示成

$$L=\{a^n b^m c^n d^m |_{n,m=1,2\cdots}\} \tag{9.4.1}$$

前面提出的建议①与建议②有时是有矛盾的，这是因为根据目前的技术水平，按建议①来选择基元会使这些基元太复杂而难以识别。而另一方面建议②强调了基元的易于抽取，而可能会使描述与分析变得很复杂。这两者之间的折衷，有时对于识别系统的实现是很重要的。例如对数学表达式作结构分析时，用字符与算符作基元就较简单，而直接选用直线曲线线段作基元就很困难，亦即在这以前应该先做一步从笔划到字符(算符)的合成操作，而不是直接地一步到位。

另一个例子是汉字识别。根据汉字结构的研究，发现可用为数不多的分割操作。例如：

分割操作：

结构关系：　左—右 $(l-r)$　　上—下 $(a-b)$　　包　围 $(S)$

每种分割操作，在相邻的两子模式或基元间产生特定的结构关系。递归地应用这些操作，我们能将汉字分割成它的子模式与基元。若用目前的技术能抽取适当的基元的话，一个汉字可用给定的一组基元与结构关系来描述，图 9-30 就是这样的例子。从这例子也可以看到，若用笔划作为基元，就会使结构描述比现在复杂得多。

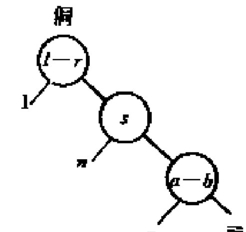



图 9-30　关于汉字“侗”的结构描述

有一篇文章报道了一种图片模式的基元分解法，通过对训练样品的观察使识别系统试探地确定基元。一个模式首先按一定顺序处理一遍，其中包括噪声滤除、空隙填充、曲线跟踪。处理的结果是产生对图片中线段的描述(直线线段用长度、斜率描述；曲线线段用长度、曲率描述)，不同线段间的结构关系用连接表来表示。总程序最后产生一个一维的符号串作为该模式的完整描述，该描述完全与

图片的设置角度、外形大小无关。前面所提到的线段长度，完全是不同线段相对而言的。

2. 模式基元的抽取

正如前面讨论所指出的，没有一成不变的办法解决基元的选择问题，因而基元抽取也只能是在基元选定了以后，才可能设计出相应的办法，下面仅以对曲线的直线段近似方法作基元抽取的示范性介绍。

在一个波形模式中，由于扫描值的系列$[f_i]$直接给出了一个曲线过程，于是，可直接通过外形元素对此曲线实行近似。相反地，在一个图像模式中，一般均需要首先实行一系列的预处理措施，使其变换为二值的线条模式。

若问题仅归结为简单模式分类，则大多数情况下，可首先借助于一个阈值算法使客体从背景中分离出来(在模式质量较差的情况下还必须在阈值运算之前执行相应的平滑去干扰运算)，从而得到一个二值图像，在客体取值1，在背景点取值0。客体的轮廓曲线则可通过下面的一个简单算法来估计：

①以行扫描方式找出一具有灰度值1的点$P$作为轮廓线的初始点。

②设想从前已找到的最后一个点出发，继续执行这一寻找过程。如该点灰度值1，则往左拐弯继续寻找；如该点灰度值0，则往右拐弯继续寻找。每一个这样找到的灰度值1的点即为轮廓线点。

③重复②，直至最后找出的轮廓线点与初始点重合。

④算法执行的结果应输出一个有序的点列，其中每一个点均在轮廓线上。

另一个方法是首先找出全部轮廓线点，即找出每一个具有灰度值1，且其具有灰度值1的邻域点小于8个的点，它们即为轮廓线点，然后再行整理输出一个轮廓线的点列。

这样的轮廓线点都构成该轮廓线的模式基元。

又如，若问题涉及到一个复杂模式的分析，则在大多数情况下，可首先对复杂模式实行一系列预处理措施，以改善模式的质量，然后采取区域分割或边缘检测的方法，检出感兴趣的客体区域。这些区域一般亦由区域的轮廓线标出。此时，轮廓线点亦可视为模式基元。

综上所述，对于模式基元的抽取，关键在于如何近似客体的边缘轮廓线。一个用来抽取模式基元以确定轮廓线的逐段线性近似方法的说明，如图9-31所示。

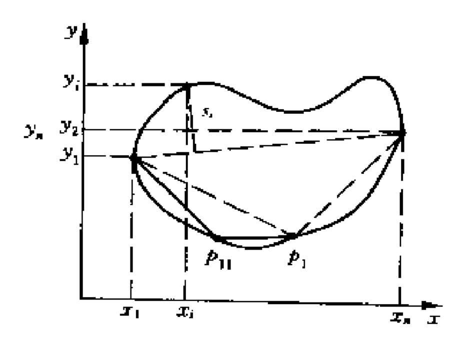


图9-31 轮廓线的逐段近似

设轮廓线的点构成一个有序点集。

$$P=\{(x_j,y_j)|_{j=1,2,\cdots,n}\} \qquad (9.4.2)$$

在一维模式$f(x)$的情况下，$x_j=j\Delta x$、$y_j=f(x_j)$。在二维模式的情况下，$x_j$、$y_j$是轮廓线上的点的坐标，而其序则给出轮廓线的一次围绕遍历中的前后串接关系，即模式基元的前后串接关系。

逐段线性近似方法的目的是，寻找一组理论上是完备的直线段作为外形元素。其基本点是，要求经过预先给定的点作出合适的直线。

因为通过$(x_j,y_j)$及$(x_k,y_k)$的一条直线，用二点式可表示其直线方程为：$(x-x_j)/(x_k-x_j)=(y-y_j)/(y_k-y_j)$，亦可改写为$(y_j-y_k)x+(x_k-x_j)y=(x_k-x_j)y_j-(y_k-y_j)x_j$，或用点斜式表示为

$$Ax+By+C=0 \qquad (9.4.3)$$

则曲线上某点 $p_i(x_i,y_i)$到上述直线的垂直距离 $S_i$ 为

$$S_i=|Ax_i+By_i+C|/\sqrt{A^2+B^2} \tag{9.4.4}$$

因而某段曲线上的点集

$$P_i=\{(x_j,y_j)\mid j=1,2,\cdots,n_i\}\subseteq P \tag{9.4.5}$$

的近似错误可选择为最大距离

$$\varepsilon_0(S)=\max_j\{S_j\} \tag{9.4.6}$$

或均方距离

$$\varepsilon_m(S)=\frac{1}{n_i}\sum_{j=1}^{n_i}S_j \tag{9.4.7}$$

为了近似式(9.4.2)定义的点集，这里给出如下简单的过程描述：

①点集 $P$ 应逐段线性地近似，其中在第一线段中，误差 $\varepsilon_0(S)$或 $\varepsilon_m(S)$均不超出一个阈值 $T$。

②第一条近似直线的初始点及终止点表示为$(x_1,y_1)$及$(x_n,y_n)$。

③若对于所有的直线，误差均不大于 $T$，则近似过程结束。否则转向操作④。

④对于直至当前状态为止已找到的具有较大误差的直线，执行操作⑤及⑥。

⑤计算直线上具有最大距离的点 $p_m$。

⑥用两条新的直线代替原直线。第一条新直线的初始点及终止点是原直线的初始点及 $p_m$ 点；第二条新直线的初始点是 $p_m$ 点及原直线的终止点。

⑦转向操作③。

因为对于一个二维模式的封闭的轮廓来说，由于$(x_1,y_1)\approx(x_n,y_n)$，故该方法并不总是有效的。然而，在许多情况下，经过下面的适当的改变仍可使用：作一任意矩形与该曲线相切，设曲线与该矩形互不相接的最远的两条边的切点为$(x_1,y_1)$及$(x_n,y_n)$，于是曲线被这两个点分为上下两条分支曲线，再按上述算法分别进行近似。近似的第一条直线可从经过点$(x_1,y_1)$及$(x_n,y_n)$的直线开始。在图 9－31 中，对于曲线下方的分支，$p_1$ 是具有最大纵坐标距离 $d$ 的点。于是，在下一步中，曲线将分别由通过$(x_1,y_1)$及 $p_1$ 的直线与通过 $p_i$ 及$(x_n,y_n)$的直线来近似。这一操作继续执行，直至达到所要求的精度为止。当然可能会出现某些直线段很短的情况，但这可通过设置阈值来避免。

### 四、模式文法

对于某一给定的应用，在解决了基元选择问题后，下一步就要建立一种(或几种)文法，以产生一种(或几种)语言，以描述待识模式。人们希望有一种具有学习功能的推理机，它能从描述待识模式的字符串集合中派生出所需要的文法，但除了某些非常专门的情况外，还无法得到这种机器。迄今为止，在大多数情况下，设计者还只能根据自己的经验来设计所需的文法，而且若要提高语言的描述能力，就会使分析系统相应地复杂起来。有限状态自动机可用于辨识或接受有限状态语言。当然有限状态语言的描述能力弱于上下文无关语言和上下文有关语言，但另一方面，处理上下文无关和上下文有关文法所产生的语言，通常要求状态无限制且非确定性的文法分析程序。故设计者应从用途出发，在某种文法的描述能力与相应语言的分析效率之间作适当折衷。

1. 串文法

串文法是一种一维文法，亦被称为链文法。

定义 9.4.1　一个串文法被定义为四元组：

$$G=(V_N,V_T,S,P) \tag{9.4.8}$$

其中有限非空集合

$$V_N=\{S,A_1,A_2,\cdots,A_n\} \tag{9.4.9}$$

为非终止符集(*nonterminal set*，有的书上亦称非终结符集)或称中间模式元集。而有限非空集合

$$V_T=\{a_1,a_2,\cdots,a_n\} \tag{9.4.10}$$

为终止符集(*terminal set*，相应地亦称终结符集)或模式基元集，且 $V_T\cap V_N=\varnothing$，$S\in V_N$ 为初始符或初始模式。有限非空集合

$$P=\{r_1,r_2,\cdots,r_n\} \tag{9.4.11}$$

称产生式集或规则集。一般地，每一产生式有如下形式

$$r_i:\alpha_i\rightarrow\beta_i,\quad (i=1,2,\cdots,n \tag{9.4.12}$$

其中

$$\alpha_i\in(V_N\cup V_T)^*V_N(V_N\cup V_T)^*$$

$$\beta_i\in(V_N\cup V_T)^*$$

在这里$(V_N\cup V_T)^*$表示 $V_N\cup V_T$ 的传递闭包，即由 $V_N\cup V_T$ 上的有限符号串组成的集合。空串用 $\lambda$ 表示。且令$(V_N\cup V_T)^+=(V_N\cup V_T)^*-\{\lambda\}$。

为表述方便，我们用大写英文字母 $A,B,C,\cdots$表示非终止符；用位在前面的 $a,b,c\cdots$等小写英文字母表示终止符；用位在后面的 $u,v,w,x,y,z$，表示终止符串；而用小写希腊字母 $\alpha,\beta,\gamma\cdots$等表示由终止符与非终止符组成的符号串。我们还用$|\alpha|$来表示符号串 $\alpha$ 的长度，用 $\alpha^n$ 表示将串 $\alpha$ 重写 $n$ 次。

由式(9.4.12)可知，每一产生式的左边即 $\alpha_i$ 中至少包含一个中间模式元，即串 $\alpha_i=\alpha_k\alpha_l\alpha_m$，其中 $\alpha_k\in(V_N\cup V_T)^*$，$\alpha_l\in V_N$，$\alpha_m\in(V_N\cup V_T)^*$。

又设 $\alpha_i,\alpha_{i+l}\in(V_N\cup V_T)^*$，如果存在子串 $\alpha_k,\alpha_l,\alpha_m,\alpha_n\in(V_N\cup V_T)^*$，使

$$\alpha_i=\alpha_k\alpha_l\alpha_m,\alpha_{i+1}=\alpha_k\alpha_n\alpha_m,\alpha_l\rightarrow\alpha_n\in P \tag{9.4.13}$$

则称串 $\alpha_{i=l}$可由串 $\alpha_i$ 直接导出，并简写为

$$\alpha_i\Rightarrow\alpha_{i+l} \tag{9.4.14}$$

如果存在串 $\alpha_i$，$i=1,2,\cdots\cdots,l$，使得

$$\alpha_m=\alpha_1\Rightarrow\alpha_2\Rightarrow\cdots\Rightarrow\alpha_l=\alpha_n \tag{9.4.15}$$

则称串 $\alpha_n$ 可由串 $\alpha_m$ 导出，并简写为

$$\alpha_m\overset{*}{\underset{G}{\Rightarrow}}\alpha_n\text{(无歧义情况下 }G\text{ 可省略)} \tag{9.4.16}$$

定义由文法 $G$ 所产生的语言 $L(G)$为

$$L(G)=\{x\mid S\overset{*}{\Rightarrow}x,x\in V_T^*\}$$

亦即 $L(G)$是借助于产生式的应用，由 $S$ 推导出来的终止符串的集合。一个终止符串 $X\in L(G)$亦称为语言 $L(G)$的一个句子。在上述终止符串推导过程中，如果每次都仅替换最左部的非终止符，则称“最左推导”。

如对由式(9.4.12)所定义的产生式加以限制，则可得到关于文法的特殊类型。

定义 9.4.2　若对产生式作如下限制：

$$\alpha_1 A\alpha_2 \to \alpha_1 \beta\alpha_2 \tag{9.4.17}$$

其中 $A\in V_N,\alpha_1,\alpha_2,\beta\in(V_N\cup V_T)^*,\beta\neq\lambda$，则此文法称为上下文有关文法。

术语"上下文有关"指的是仅当非终止符 $A$ 出现在子串 $\alpha_1,\alpha_2$ 的上下文之间时，才能被重写为 $\beta$。

定义 9.4.3　若对产生式作如下限制：

$$A\to\alpha \tag{9.4.18}$$

其中 $A\in V_N,\alpha\in(V_N\cup V_T)^+$，则此文法称为上下文无关文法。即允许非终止符 $A$ 被串 $\alpha$ 所替换，而与 $A$ 的上下文无关。

定义 9.4.4　若对产生式的限制为

$$A\to aB \quad 或 \quad A\to a \tag{9.4.19}$$

其中 $A,B\in V_N,a\in V_T$，则此文法称为正则文法或有限状态文法。

定义 9.4.5　对产生式不作限制的文法称为无约束文法。

不同的文法可产生不同的语言，上下文有关文法、上下文无关文法及正则文法分别产生上下文有关语言、上下无关语言及正则语言。而且从上面对这三种文法的定义不难看出：

正则语言⊂上下文无关语言⊂上下文有关语言。

例如：$L_1=\{a^nb^nc^nd^n|_{n=1,2,\cdots}\}$是上下文有关语言，而非上下文无关的；$L_2=\{a^nb^n|_{n=1,2,\cdots}\}$是上下文有无关语言，而非正则的；$L_3=\{a^nb|_{n=1,2,\cdots}\}$是正则语言。

上下文无关文法可以产生正则文法所能产生的语言 $L_3$，但正则文法不能产生上下无关文法所能产生的全部语言 $L_2$，这一结论同样适用于 $L_2,L_1$。

对模式识别问题而言，在理论和实践上都极为重要的问题是，哪一类型的语言是"可决定的"(句子问题的可决定性)，即决定一个字符串是否是这个语言的一个句子。换句话说，即决定一个给定的模式是否属于由产生该语言的文法所定义的模式类。显然，在模式识别问题中，人们感兴趣的仅是那些具有句子问题可决定性质的语言。在这里，可指出一个在形式语言理论中的相应结论，对于一个句子 $x\in V_T^*$，一个由无约束文法产生的语言，不能对它作出上述决定。也就是说，在模式识别中，只有上下文有关文法、上下文无关文法及正则文法才是可利用的，因为它们在一个有穷的运算步骤内，可以判定一个模式是否属于一个确定的类别。而且，在模式识别中，经常采用的是正则文法和上下文无关文法，因为在这两个文法中，判定一个句子是否可以由相应的文法产生，是特别简单的。

下面以一些简单的例子，说明串文法怎样用来描述模式。如描述边长为 $n$ 个长度单位的等边三角形、矩形以及描述一个染色体等。

例 9.4.2　建立一种文法以产生有限状态语言 $L=\{a^nb^nc^n|_{1\le n\le 3}\}$，可以看出这种语言可用来描述边长为 $n$ 个单位长度的三角形，为使该种文法与一自顶向底，面向目标的过程段相兼容，该种语法应严格按照自左向右的顺序产生终止符。在大多数情况下，任一产生式的一次运用就会产生一个终止符。非终止符不得出现在终止符的左面。

(1) 以正则文法实现此语言 $L$。

$$G_1=(V_N,V_T,P,S)$$

其中 $V_N=\{S,[B^i,C^j]\,|\,0\le i,j\le 3\},V_T=\{a,b,c\}$

$P$：

1. $S \to a[B,C]$,　　2. $[B,C] \to a[B^2,C^2]$,
3. $[B^2,C^2] \to a[B^3,C^3]$,　　4. $[B,C] \to b[C]$,
5. $[C] \to c$,　　6. $[B^2,C^2] \to b[B,C^2]$,
7. $[B,C^2] \to b[C^2]$,　　8. $[C^2] \to c[C]$,
9. $[B^3,C^3] \to b[B^2,C^3]$,　　10. $[B^2,C^3] \to b[B,C^3]$,
11. $[B,C^3] \to b[C^3]$,　　12. $[C^3] \to c[C^2]$。

(2) 以上下文无关文法获得此语言 $L$。

$$G_2 = (\{S,B,C\},\{a,b,c\},P,S)$$

其中 $P$:

1. $S \to aSBC$,　　2. $S \to abC$,
3. $CB \to BC$,　　4. $bB \to bb$,
5. $bC \to bc$,　　6. $cC \to cc$。

(3) 以上下文无关程序文法获得此语言 $L$。

在本文导出过程中,产生式用于中间链,下一次选用的产生要从本次所用产生式的成功区中选择。如果成功中的产生式不能用,则从失败区中选择产生区。成功区与失败区合称特征区。如果可用的特征区仅包含∅,则导出过程停止。

在此我们给出以下两种等效的文法 $G_3, G_4$:

$G_3 = (V_N, V_T, J, P, S)$

其中　$V_N = \{S,A,B,C\}$,　$V_T = \{a,b,c\}$,

$J = \{1,2,3,4,5,6,7\}$。

$P$:

| 标号 | 核 | 成功区 | 失败区 |
|---|---|---|---|
| 1 | $S \to ABC$ | {2,5} | {∅} |
| 2 | $A \to aA$ | {3} | {∅} |
| 3 | $B \to bB$ | {4} | {∅} |
| 4 | $C \to cC$ | {2,5} | {∅} |
| 5 | $A \to a$ | {6} | {∅} |
| 6 | $B \to b$ | {7} | {∅} |
| 7 | $C \to c$ | {∅} | {∅} |

$G_4 = (\{S,A,B,C\},\{a,b,c\},\{1,2,3,4,5,6,7\},P,S)$

其中 $P$:

| 标号 | 核 | 成功区 | 失败区 |
|---|---|---|---|
| 1 | $S \to ABC$ | {2,5} | {∅} |
| 2 | $A \to aA$ | {∅} | {3} |
| 3 | $B \to bB$ | {∅} | {4} |
| 4 | $C \to cC$ | {2,5} | {∅} |
| 5 | $A \to a$ | {∅} | {6} |
| 6 | $B \to b$ | {∅} | {7} |
| 7 | $C \to c$ | {∅} | {∅} |

即使从这一简单例子,我们也可以看出,上下文无关文法的能力较正则文法强得多,而程

序文法则有利于程序控制流程建立。

2. 阵列文法

本小节介绍的阵列文法是直接描述二维或多维关系的另一种方法。

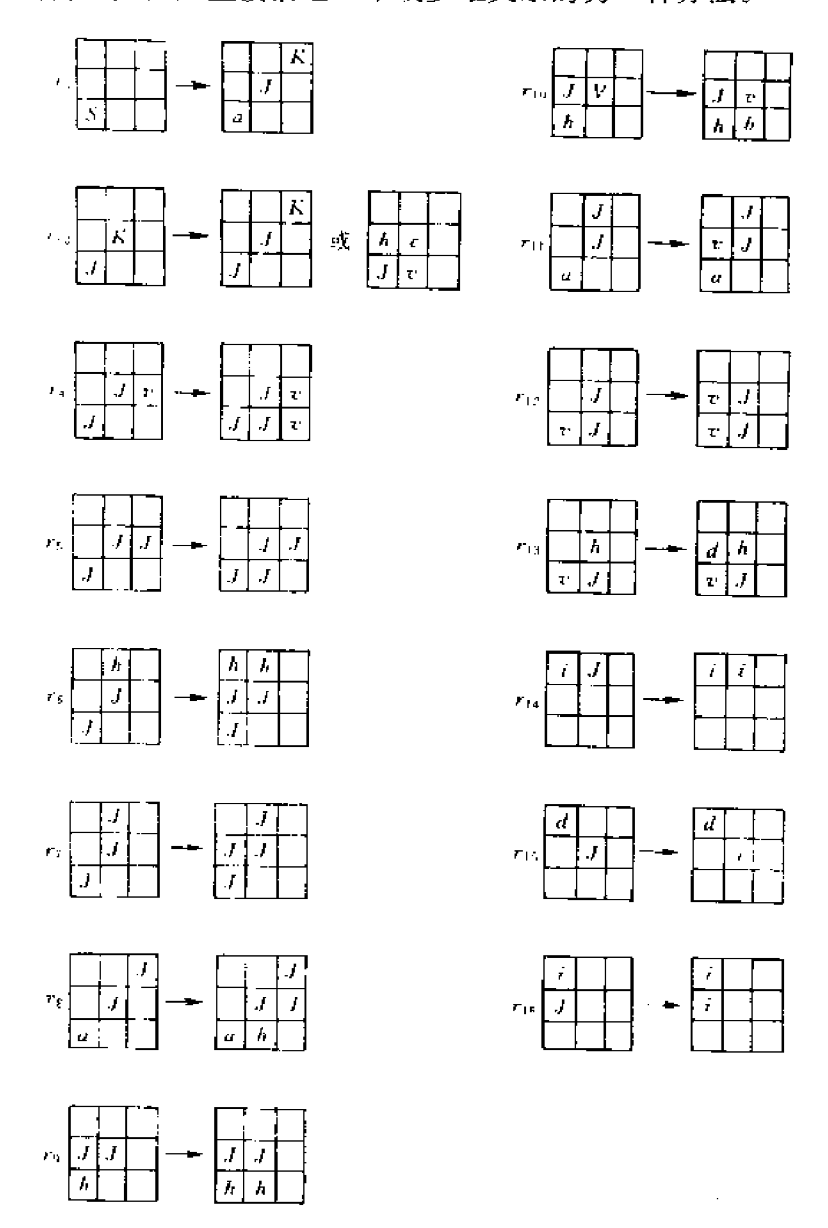


图 9-32 阵列文法一例

例 9.4.3 为了与链文法进行比较，首先给出一个阵列文法的例子。该文法同样产生一个具有单位长度整数倍边长的正方形。以 $a$、$b$、$c$、$d$ 表示正方形的直角点，$h$、$v$ 表示水平及垂直方向的线段，$i$ 表示内点。可以看出，所给出的文法，可以直接利用模式的二维关系。

文法 $G_1=(V_N,V_T S,P)$

其中：$V_N=\{S,J,K\}$

$V_T=\{a,b,c,d,h,v,i\}$

$P=\{r_1,r_2,\cdots,r_{16}\}$

$r_1 \sim r_{16}$在图 9-32 中给出。

$S_0$

| d | h | h | h | c |
|---|---|---|---|---|
| v | i |  | i | v |
| v | i | i | i | v |
| v | i | i | i | v |
| a | h | h | h | b |

图 9-33 一个正方形的二维描述

运用上述文法，可得到一个关于所述正方形的二维描述，如图 9-33 所示。

定义 9.4.3　一个阵列文法形式地定义为一个四元组

$$G=(V_N,V_T,P,S)$$

其中 $P$ 是二维产生式的集合，每一产生式有如下的形式：

$$F_{i1}\rightarrow F_{i2}$$

其中 $F_{i1}$，$F_{i2}$ 均为由终止符与非终止符组成的二维阵列(参见图 9-32)。

在二维阵列替换中若 $F_{i2}$ 的行列数大于 $F_{i1}$，则将导致模式的破坏，为避免此种情况，总是取 $F_{i1}$ 与 $F_{i2}$ 的阶数相同。但另一方面符号阵列在替代过程中的增长又是必需的，故引入空格并以_表示_$\overline{\in} V_N\cup V_T$，于是一个阵列文法的产生式可形如

$$\begin{bmatrix}\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n}\\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n}\\ \vdots & \vdots & & \vdots\\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn}\end{bmatrix}\longrightarrow\begin{bmatrix}\beta_{11} & \beta_{12} & \cdots & \beta_{1n}\\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n}\\ \vdots & \vdots & & \vdots\\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn}\end{bmatrix}$$

其中 $\alpha_{ij},\beta_{ij}\in V_N\cup V_T\cup\{_\}$，$1\leqslant i\leqslant m$，$1\leqslant j\leqslant n$ 且有

(1) 若 $\alpha_{ij}\neq_$，则 $\beta_{ij}\neq_$；

(2) 若 $\alpha_{ij}\in V_T$，则 $\beta_{ij}=\alpha_{ij}$；

这样两条规则。

以上空格的引入实现了一个阵列的增长。初始阵列含有初始模式，它全部由空格组成，因为空格允许由来自 $V_N\cup V_T$ 的符号所替代(相反则不然)，故阵列在替代过程中实际上是增长的。

## 五、串的识别与分析

对两类模式进行分类，应判别描述该模式的字符串 $x$ 是否可由文法 $G$ 来产生。若 $G$ 能生成，则 $x\in W_1$，反之 $x\in W_2$。推广到 $M$ 类分类，应先确定 $M$ 种文法，它能生成 $M$ 类语言 $L(G_i)$，$i=1,2,\cdots,M$。一个描述待识模式的字符串 $x$，当它是属于 $L(G_i)$ 中的一个句子，就应属于 $W_i$ 类。串识别器的原理框图如 9-34 所示。

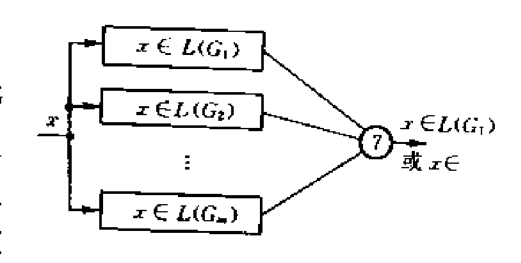



图 9-34　串识别器原理图

用自动机技术识别未知模式的方式，本质上是对输入字符串从左向右地检查，以判断它能否为机器所接受。如果 $x$ 是由正则文法产生的，则可构造一个确定的有限状态自动机识别之；如果 $x$ 是由上下文无关文法产生的，则相应的识别器通常是一个非确定下推自动机。由于并非每一非确定下推自动机均能有一等价的确定下推自动机，而前面已指出上下文无关文法是描述模式结构的重要方法，故除了自动机技术外，还提出了许多上下文无关语言的句法分析算法，如算子优先法、Cyk 算法、Earley 算法、转移图法等。

用自动机技术，计算量较小，但它仅能实现串的识别而没有对语法结构进行分析，然而在有些场合不仅需要判断字符串是否属于 $L(G)$，同时还要知道它的句法结构，以便改进文法。这就要采用句法分析器(或剖析)。

句法分析器的输出通常不仅包括对由给定文法所产生的串的识别，而且有的文法还能给出串的导出树，进而给出模式的完整结构描述。

从方法上来看，Cyk 算法是一个自下而上的句法分析算法。Earley 算法本质上是自上而下

的句法分析算法。转移图法也是自上而下算法，它与 Earley 算法有很多相似之处，它包含了 Earley 算法三种运算中的两种：预测运算与完成运算。由于缺少扫描运算，所以对上下文无关文法的形式有一定要求，这就使它应用起来似乎不及 Earley 算法灵活。

Cyk 算法概念清晰，但要求产生式是乔姆斯基范式化了的，这一限制使它在文法推断上不很方便，而且乔姆斯基范式化后，产生式和非终结符数量都显著增加，使分析搜索时间增加，所以它的运算比 Earley 算法长。

一个句子能否被某种文法所接受，Earley 算法很容易判别，往往不必将运算进行到底，即可判别，而 Cyk 算法则一定要到全部计算完毕，才能判定一个句子能否被某文法所接受，所以在多类判别时，Earley 算法有更大优越性。

## §9.5　神经网络及其在模式识别中的应用

### 一、人工神经网络概述

1. 基本模型

早在 50 年代，研究人员就开始模拟动物神经系统的某些功能，他们采用软件或硬件的方法，建立了许多以大量处理单元为结点，处理单元间实现（加权值的）互联的拓扑网络，进行上述模拟，并称之为人工神经网络。不言而喻，人工神经网络中的处理单元是人类大脑神经元的简化，处理单元间的互联则是轴突—树突这一信息传递路径的简化，这种模拟确实在某种程度上接近人类思维的部分机理，故在切实的算法诞生后，便出现令人鼓舞的成功。

一个处理单元（现在已当作一个人工神经元）将接受的信息 $x_0, x_1, \cdots, x_{n-1}$，通过用 $W_0, W_1, \cdots, W_{n-1}$ 表示的互联强度，以点积的形式合成为自己的输入（见图 9-35），并将输入与以某种方式设定的阈值 $\theta$ 作比较，再经某种形式的作图函数 $f$ 的转移，便得到该处理单元的输出 $y$。常用的三种非线性作用函数 $f$ 的形状，如图 9-36 所示。其中 9-36(c)图所示的作用函数，被称为 Sigmoid 型，简称为 S 型，是极常用的。

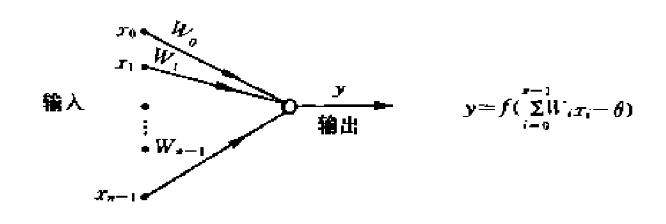



图 9-35　一个人工神经元的输入及输出

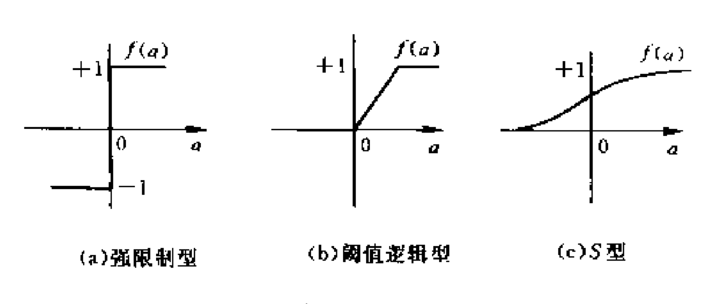



图 9-36　常用的作用函数曲线

处理单元的输入与输出间的关系由下式给出

$$y = f\left(\sum_{i=0}^{n-1} W_i x_i - \theta\right) \tag{9.5.1}$$

式中：$x_i$ 为第 $i$ 个输入元素（通常为 $n$ 维输入矢量 $X$ 的第 $i$ 个分量）；$W_i$ 为从第 $i$ 个输入与处理单元间的互联权重；$\theta$ 为处理单元的内部阈值；$y$ 为处理单元的输出。

上面我们已给出了一个有代表性的处理单元或人工神经元的工作过程。出于不同的使用

目的，现在已研制出为数众多的神经网络模型及表征该模型动态过程的算法，如反向传播(BP)算法、Hopfield 算法等等以及它们的改型。从图 9－36 中可以看出在人工神经元中通常采用着非线性的作用函数，当大量神经元连成一个网络并动态运行时，则构成一个非线性动力学系统，虽然单个神经元的工作过程较简单，但整个系统是非常复杂的，它具备一般非线性动力学系统的全部特点，如不可预测性、不可逆性、多吸引子等，故前面所说的算法，通常也只是刻划单个神经元动力学过程的描述而很难是全系统的描述，这一点是值得提醒读者的。这样一个复杂的非线性动力学系统作为对人脑的模拟，呈现出高维性(一个系统有众多的神经元)、自组织性、模糊性(某种程度的"表决")、冗余性(部分处理单元的错误不影响整个问题的解)等优良品质与可贵的自学习能力(传统计算机一般不具备这种能力)，而且较冯·诺以曼体系更适合于人类大脑思维机理的模拟。但是应该清醒地看到这种模拟还是极肤浅的，这一方面是由于人类对自身思维机理的认识尤其肤浅，另一方面也由于现实的可行性原因而对人工神经元作了极度的简化，从而也影响了思维模拟的效果。

尽管如此，由于业已展出的一系列优良品质，故已在传感器与信号处理、知识工程、最优化问题求解、过程建模与控制等方面展示了很好的实用效果与应用前景。

### 二、与传统模式分类器的对比

现在介绍 6 种重要的用于模式分类的神经网络，并以此与传统模式分类器进行对比(图 9－38)。与传统分类问题一样，现在假设它们均是用于将 $n$ 维的样品分类归属 $m$ 类模式中之某一类。

图 9－37 上部示出的传统分类器包含两级：第一级计算待识样品(以下简称样品)对各类模式标本(以下简称样本)的匹配程度；第二级选出具有最大匹配度的类别。第一级的输入是用符号表示的 $n$ 个输入元素的值，它们顺序地译码为有利于运算的内部形式。然后要设计出一种算法，以算出样品与每一类样本的匹配程度。显然每一样本应是该类模式的代表，而样品则往往是由样本以某种随机方式产生的。在这种情况下，总是假设样品的分布具有某种函数形式(例如正态分布函数等)，因为这使匹配度的计算较简单。然后匹配度被顺序加载到分类器的第二级，并选出具有最大匹配度的类。

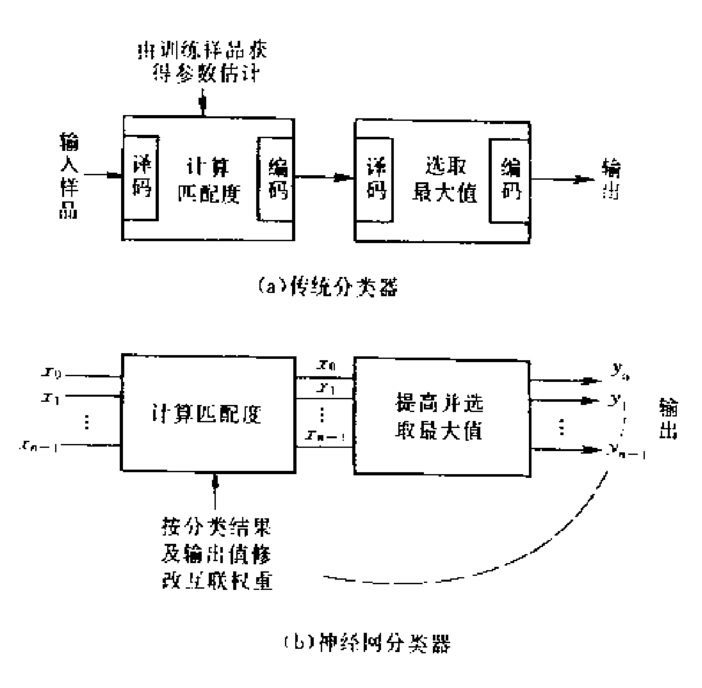



图 9－37　传统分类器与神经网分类器的对比

图 9－37 的下部则是神经网络分类器，用 $n$ 个分量表示的样品被送入神经网络，这些分量可用二值表示，或用连续值表示。神经网络的第一级实际上也是在计算匹配度，然后被平行地通过 $m$ 条输出线送到第二级，在第二级中各类均有一个输出，并表现为仅有一个输出的强度为"高"，而其余的均为"低"。当得到正确的分类结果后，分类器的输出可反馈到分类器的第一

级，并用一种学习算法修正权重。当后续的测试样品与曾学习过的样本十分相似时，分类器就会作出正确的响应。

图 9－38 中给出这 6 种用于分类的神经网络的树状表示。第一层分枝是根据输入值为二值变量还是连续变量；第二层分枝则是根据训练是否有监督。诸如 Hopfield 网，感知器网等是有监督训练的网，传统分类器（例如正态分类器等）都是有监督的，即所用的训练样本是已知类别的，图 9－38 中第二层的 Kohonen 网等则是无监督的。网络间的差异还在于是否支持自适应的训练，不过这一点在图 9－38 中未能体现出来。

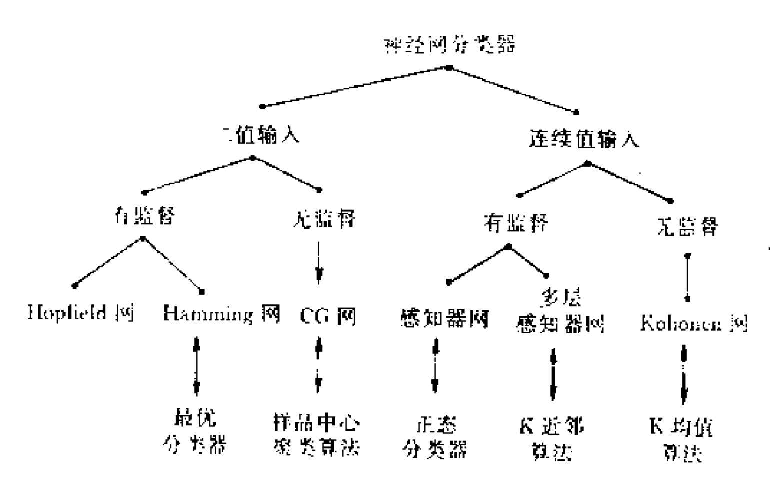



图 9－38　6 种神经网分类器的分类树

图 9－38 下部还给出了与网络相对应的经典算法，这种对应关系有时是很好的。例如 Hamming 网确实与用于含有噪声的二值模式的最优分类器很相似。但有时却并非如此，例如感知器网与正态分类器的特性并不相同；Kohonen 网也并不执行迭代 K 均值训练算法，而是每一新模式加载后，权值就作相应修改。然而 Kohonen 网与 K 均值算法，在聚类数 K 的预先指定上则是一致的。

1. Hopfield 网

从图 9－38 的树状分层中已可看出，它通常是用于二值输入模式的，例如黑白二值表示的象素图形。Hopfield 网在联想存贮器及求最优解等问题上的能力是很好的。

(a) 8 个样本模式

(b) 含有噪声的输入"3"及其输出

图 9－39　Hopfield 网应用之一例

图 9－39(a) 为 Hopfield 网的应用之一，该网有 120 个结点，有 14400 条互联，使用图在图 9－39(b) 所示出的例子中，数字"3"的样本以 0.25 的概率被污染，即将原样本以 0.25 的概率使黑白象素颠倒，以构成待识样品，该样品加载入网络并经 7 次迭代后所得之输出位于图 9－39(b) 下部。

2. Hamming 网

前面已提到在观察 Hopfield 网的能力时，用到将样本中部分象素黑白颠倒的办法，这实际上是通信理论中的一个经典做法。最小误差分类器这时要计算样品与每一样本类的 Hamming 距离，它实际上是样品与相应样本不一致的位数（bits）。这样的网络亦被称为 Hamming 网，并示于图 9－40。

算法 9.5.1（Hamming 网算法）

步骤 1：指定互联权重及阈值。

在下层子网中：

$$W_{ij}=x_i^j/2 \qquad \theta_j=n/2$$

其中　$0\leqslant i\leqslant n-1, 0\leqslant j\leqslant m-1$

在上层子网中：

$$T_{kl}=\begin{cases}1, & k=l\\ -\varepsilon, & k\neq 1,\varepsilon<1/m\end{cases}$$

其中 $0\leqslant k,l\leqslant m-1$

在上述公式中，$W_{ij}$为输入 $i$ 到下层子网的结点 $j$ 的互联权重，$\theta$ 为这些结点的阈值；上层子网中结点 $k$ 到结点 $l$ 的互联权重为 $T_{kl}$，阈值则全为 0. $x_i^j$ 是样本 $j$ 的第 $i$ 个元素。

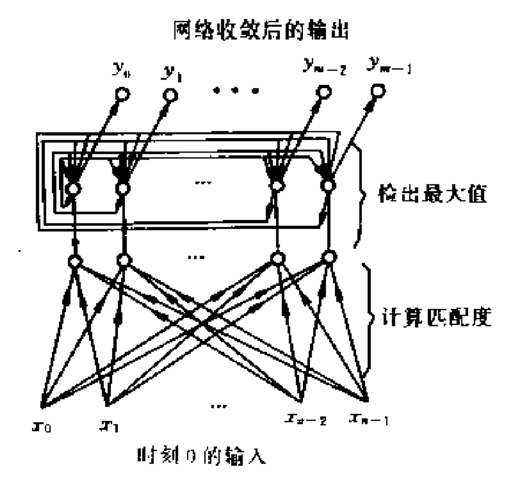



图 9-40　前馈 Hamming 网最大似然分类器

步骤 2：用未知的输入模式初始化。

$$y_j(0)=f\left(\sum_{i=0}^{n-1}W_{ij}x_i-\theta_j\right)$$

本算法中 $y_j(t)$为输出结点 $j$ 在 $t$ 时刻的输出，$x_i$ 为输入的第 $i$ 个元素，$f$ 为非线性阈值函数。在此处以及在以后均假设非线性函数的最大输入均不致使输出饱和。

步骤 3：迭代直到收敛。

$$y_j(t+1)=f\left(y_j(t)-\varepsilon\sum_{k\neq j}y_k(t)\right)\quad 0\leqslant j,k\leqslant m-1$$

这一步骤不断重复直到收敛，即此后仅一个结点的输出为正。

由此算法可以看出，在下层子网中首先将权值与阈值设置好，这样，由图 9-41 中部结点所产生的输出就是匹配度，其值为样品维数 $n$ 减去到每一样本的 Hamming 距离，故取值范围在 0 与 $n$ 之间，具有最大值的结点对应着与输入匹配最好的样本。Hamming 网的上层子网的阈值与互联权重是固定的。全部阈值均置为 0，从每一结点到它自身的互联为 1，结点间是抑制的，故互联为$-\varepsilon$，其中 $\varepsilon\leqslant 1/m$。

在权重与阈值设定后，便可加入具有 $n$ 个分量的二值待识样品，该样品应有足够长的加载时间，以完成下层子网中的匹配度计算，并使上层子网的输出稳定在某一初值上；这时可撤消输入，上层子网中将不断迭代直到仅有一个结点的输出为正，该结点对应的模式即为识别结果。

图 9-41 为 Hamming 应用之一例。4 个分图分别是 100 个输出结点在 0,3,6,9 次迭代后的结果，在该实例中样品的维数高达 1000，样品是属于第 50 类的，当在时刻 0 时，首先加载到网络上，然后将它撤消，这时第 50 号结点的输出最大(为 1000)，而其它结点为 500 左右的随机值。在 3 次迭代后，除第 50 号结点外，其它结点的输出均大大下降，而在 9 次迭代后仅 50 号结点的输出大于 0，模拟结果表明，以不同概率使输入样品的 0、1 位颠倒，或改变输入样品的类别数与维数，一般在 10 次迭代后，均能收敛。模拟结果还表明，取 $\varepsilon<1/m$ 时，输出结点能达到最大值。

较之 Hopfield 网而言，Hamming 网所需要的互联较少。例如识别 100 个分量 10 类样品的 Hamming 网，仅需 1100 个互联，而 Hopfield 网则需 10000 个互联。而且当分量数与类别数增加时，这种优点就更明显。因为在 Hopfield 网中，互联数随样品维数按平方关系增长，而 Hamming 网仅是线性地增大。

在模式分类中经常遇到要求选出最大输入值的问题，Hamming 网中可采用很强的侧向抑制的办法，构成一个“赢家独吞”的网络。图 9-42 中采用比较子网络检测最大值的办法，它在用阈值逻辑测出两个输入中较大者后，将此值不作修改地向上传递。一般说来用 $\log_2 m$ 层比较

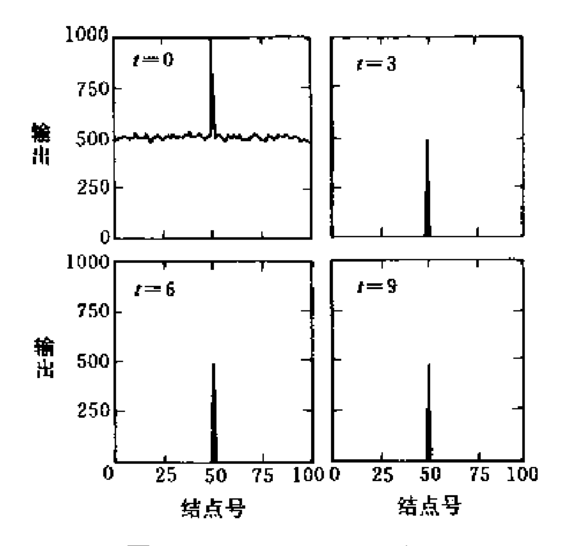



图 9-41 Hamming 网应用一例

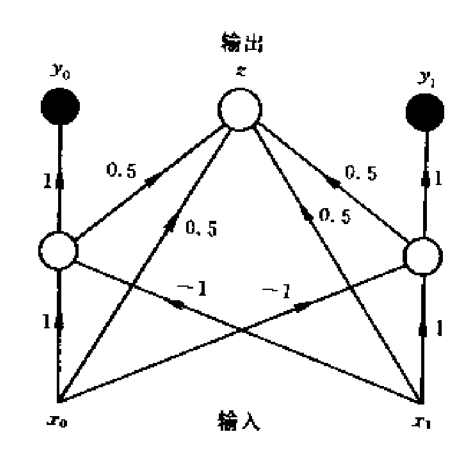



图 9-42 选择最大输入的比较网络

子网络即可检测出 $m$ 个输入中之最大者。

## 三、BP 模型及其在模式识别上的应用

### 1.BP 模型的背景

50 年代末就提出了两种神经网络的模型，它们是感知机——Perceptron(Rosenblatt,1959,1962)与最小均方联想机——LMS(Windrown 和 Hoff,1960)。前者用线性阈值单元作输出单元，后者用纯性单元作输出单元。有关这两种模型的重要定理均已被证明，并被用于模式分类中。

感知机通过算法的学习过程，确能找到一组实现正确分类的权重，但前提是该两类模式应是线性可分的，否则算法将不收敛，例如二维空间中的导或问题就是一个典型的例子。

但是对于异或函数，通过增加一维(即加入适当的新特性)就变为可解问题了。在图 9-43 中，前二维用作表示原来的异或问题，第三维是前二维之“与”，即仅当前二维均为 1 时，第三维才为 1，否则为 0。

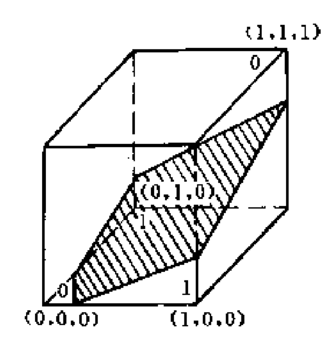



图 9-43 异或问题的三维解

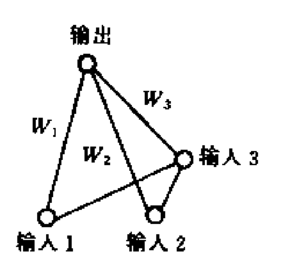



图 9-44 将二维异或问题变换为三维线性可分问题的多层网

在图 9-43 所示的立方体中存在一系列与阴影线所示平面相平行的解平面，在此解平面之上划分为一类(图中表示为 0 类)，在此解平面之下，则划分为另一类(图中表示为 1 类)。对应于这种三维问题的网络，则应为图 9-44 所示的三层网，因为结点输入 3 接受输入 1，输入 2 这两结点的输出，同时向输出结点输出自己的活跃值。

正如 Minsky 和 Papert 所判断的那样，总能把任何一问题变换为多层感知器网中可解的问题。多层网是在输入层与输出层之间增加一层或几层隐含层，在隐含层中包含着新的特性，因而就要求新的学习算法，这时自然地联想到 Widrow 与 Hoff(1960)提出的最小均方(Least-mean Square)学习算法，即 LMS 算法。

在 LMS 算法中使用线性单元，输出单元 $i$ 的活跃值 $y_i$，简单地由 $y_i=\sum_j W_{ij}x_j$ 算出，而其误差函数则为方差之和。总误差 $E$ 被定义为

$$E=\sum_p E_p=\sum_p\sum_i(t_i^p-y_i^p)^2 \tag{9.5.2}$$

其中，$p$ 为一组输入模式；$t_i^p$ 表示在输入模式为 $p$ 时，输出结点 $i$ 的期望值；$y_i^p$ 表示在输入模式为 $p$ 时，输出结点 $i$ 的实际值。

我们的目的则是训练网络以找到一组权重，使该函数极小化。

图 9-45 中给出总误差与网络中单个权重之间典型的关系曲线。

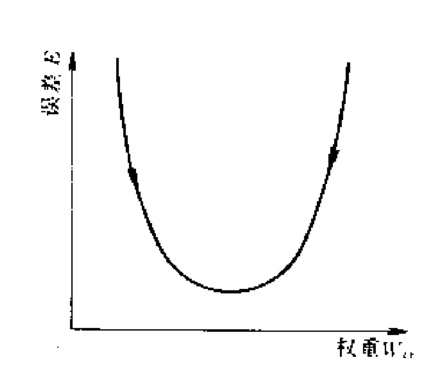


图 9-45 权重与误差的关系曲线

LMS 算法用到梯度下降法，即权重的增量正比于误差的负导数，即

$$\Delta W_{ij}=-k\frac{\partial E_p}{\partial W_{ij}} \tag{9.5.3}$$

式中 $k$ 为比例系数。

图 9-45 已告诉我们采用负导数的原由，因为当权重过大时，误差曲线在该点的斜率为正，就必须减小权重；反之当权重过小时，斜率为负，负导数的选用就加大权重。

对式(9.5.2)求导数，可解得

$$\Delta W_{ij}=2k(t_i^p-y_i^p)x_j^p \tag{9.5.4}$$

这样就将权重的变化与 $p$ 模式下输出单元 $i$ 的期望值与实际值之差联系起来了，即权重的变化正比于期望值与实际值之差乘以对应的输入。若按式(9.5.4)修改每一权重，则它们将移向各自的误差极小值，从而整个系统表现出在权重空间中向误差极小下移的特性。

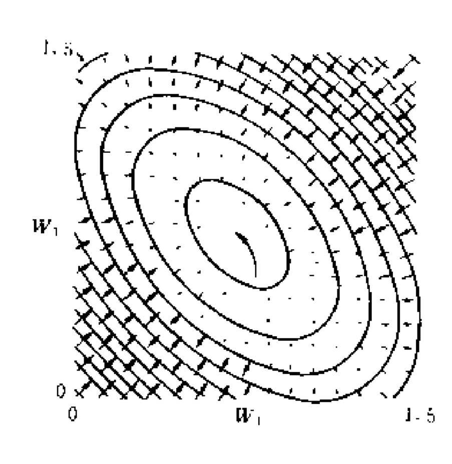


图 9-46 一个简单模型的误差等值线图

为对这一过程有更深的理解，最好能研究整个误差空间，然而描述高维空间的图形是十分困难的，故仅用图 9-46 给出只有两个权重的网络的误差下移情况，但这并不失去一般性。被考虑的网络只有两个输入结点和一个输出结点(故只需两个权重 $W_1$、$W_2$)，而且不设偏置项。用这样一个简单网络来解决“异或”问题，即输入输出对为：00→0，01→1，10→1，11→1。对应于这样的输入输出对，可通过简单计算表明在 $W_1=W_2=0.75$ 时，得到最小均方误差。在图 9-46 中等误差的轮廓线呈椭圆形，椭圆形周围的箭头表示两个权重在这些点上的导数，即表示在误差平面的这一点上两个权重的变化方向和变化的量值，可以看出图 9-46 像一个椭圆形的盘子，底部平坦，边缘较陡(请联系图 9-45 一起思考)。图中长而弯曲箭头表示从远离误差极小的一点开始，逐渐下移到实际极小的轨迹，

这一权重的轨迹当然与各该点上权重变化的方向相吻合，因而必然与权重等值线垂直。

该图说明梯度下降学习算法的一个重要事实，即它使对误差极小化起重要作用的参数发生尽可能大的变化，实际上就是使表征各权重变化合量的这一矢量指向该误差下降最快的方向。

2. BP 模型的算法及其特点

BP 算法是 LMS 算法的一般化，事实上 BP 模型正是将非线性多层感知器系统的判决能力与 LMS 算法使均方误差函数梯度极小化相结合的产物。图 9－36(c)中给出的 Sigmoid 型函数恰好能满足非线性与连续可微的条件，故 BP 算法中大多选用 Sigmoid 型函数作为输出函数 $f(x)$，即

$$f(x)=\frac{1}{1+e^{-(x-\theta)}} \tag{9.5.5}$$

式中 $\theta$ 为阈值亦常被称作偏置项。

由于已交代清楚了 BP 网是多层感知器网与 LMS 算法相结合的产物，我们在 BP 模型的背景讨论的基础上就可以直接给出 BP 学习算法，然后再讨论该种网络的一些特性。

(1)BP 学习算法：

算法 9.5.2(BP 算法)

步骤 1：将全部权值与结点的阈值预置为一个小的随机值。

步骤 2：加载输入与输出。

在 $n$ 个输入结点上加载 $n$ 维输入向量 $X$，并指定每一输出结点的期望值 $t_i$。若该网络用于实现 $m$ 种模式的分类器，则除了表征与输入相对应模式类的输出结点期望值为 1 外，其余输出结点的期望值均应指定为 0。每次训练可从样本集中选取新的同类或异类样本，直到权值对各类样本均达到稳定。实际上为保证好的分类效果，准备足够数量的各类样本，常常是必需的。

步骤 3：计算实际输出 $y_1,y_2,\cdots,y_m$。

现在是假设欲将 $m$ 类模式分类，故应按式(9.5.5)计算各输出结点 $i(i=1,2,\cdots,m)$的实际输出 $y_i$。

步骤 4：修正权值。

权值修正是采用了 LMS 算法的思想，其过程是从输出结点开始，反向地向第一隐含层(亦即存在多层隐含层时最接近输入层的隐含层)传播由总误差诱发的权值修正，这也是“反向传播”(Back-Propagation)这一称谓的由来。下一时刻的互联权值 $W_{ij}(t+1)$由下式给出

$$W_{ij}(t+1)=W_{ij}(t)+\eta\delta_j x'_i \tag{9.5.6}$$

式中，$j$ 为本结点的序号，$i$ 则是隐含层或输入层结点的序号；$x'_i$ 或者是结点 $i$ 的输出或者是外部输入；$\eta$ 为增益项；$\delta_j$ 为误差项，其取值有两种情况：

①若 $j$ 为输出结点则

$$\delta_j=y_j(1-y_j)(t_j-y_j) \tag{9.5.7}$$

其中，$t_j$ 为输出结点 $j$ 的期望值；　$y_j$ 为该结点的实际输出值。

②若 $j$ 为内部隐含结点则

$$\delta_j=x'_j(1-x'_j)\sum_k\delta_k W_{jk}$$

其中 $k$ 为 $j$ 结点所在层之上各层的全部结点。

内部结点的阈值以相似的方式修正，即把它们设想为从辅助的恒定值输入所得到的互

联权。

另外，若加入动量项，则往往能使收敛加快，并使权值的变化平滑。这时 $W_{ij}(t+1)$ 由下式给出

$$W_{ij}(t+1)=W_{ij}(t)+\eta\delta_j x_i'+\alpha(W_{ij}(t)-W_{ij}(t-1)) \qquad (9.5.8)$$

其中 $0<\alpha<1$。

步骤 5：在达到预定误差精度或循环次数后退出，否则转步骤 2 重复。

(2)BP 模型的特点：

①局部极小。BP 算法继承了 LMS 算法的误差梯度下降以致达到极小的思想。这对由图 9-46 所表示的两互联线性模型而言，是易于实现的，然而在多层网络模型中，存在着许多极小，这不仅使误差空间的表示变得困难，而且往往由于停留在局部极小中而无法得到全局最优解，所以这是对 BP 算法的研究与改进的重要方面。

为讨论局部极小，不妨设计一个三结点网络，即输入、隐含、输出结点均各一个的三层网络，它的互联权也仅两条即 $W_1$ 是输入结点与隐含结点间的互联权，$W_2$ 是隐含结点与输出结点间的互联权。该网络在阈值被置成 0 时的误差空间示于图 9-47。

这个由三结点三层网所形成的误差等值线图像一个马鞍，其左上或右下权重变化较大，当从两侧向中部靠拢后就分别向左下和右上移动，并在那里形成权重几乎不变的区域即极小区，这样就有了两个极小区。图 9-47 中的两条长弧线箭头分别表示从不同的初始权值开始，系统可能到达的两个最终状态，而且系统一旦进入其中之一就无法再跳出来，如该极小恰好不是全局极小，那么就意味着系统未能得到最优解。但研究表明，在 *BP* 网络中采用较多的隐单元，会有效地减少局部极小的个数。

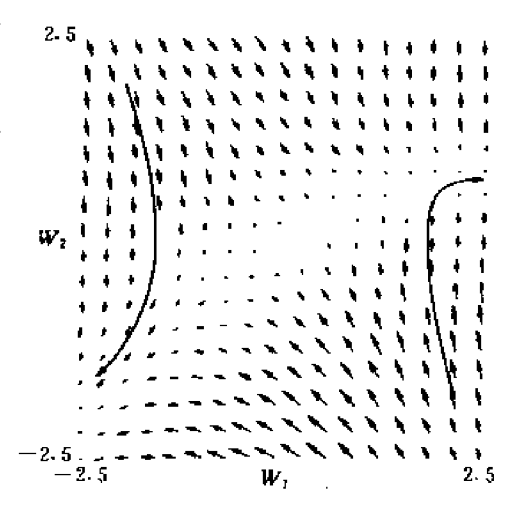


图 9-47　一个简单的三结点三层网络的误有效期等值线图

②动量。在学习过程中应使权重按误差的导数成正比例地修正，式(9.5.6)中的增益项 $\eta$ 反映了这种修正的速率。$\eta$ 太小，则学习的效率太低；反之若 $\eta$ 太大，则可能引起振荡。为此在式(9.5.8)中引入动量项，即权重的变化不仅与 $\eta$ 有关，而且与 $W_{ij}(t)-W_{ij}(t-1)$ 有关，这样就可以滤除权重空间中误差曲面的高频偏差(即误差曲面中高曲率的剧变)，而使有效的权重间隔加大(即学习次数减少)。式(9.5.8)中常用的 $\alpha$ 值为 0.8。

(3)应用实例：

由于 BP 模型将多层感知器网络与 LMS 算法相结合，通过不断比较网络的实际输出与指定期望输出间的差异来不断地调整权值，直到达全局(或局部)极小值，不难想象此种网络对模式识别中的诸多问题，均有良好的效果，事实上，在语音、文字、图像等领域已有大量成功地应用这种网络进行模式识别的报导，限于篇幅，本节仅介绍两个实例。

①二维图像的识别[9-3]：

图 9-48 给出 F－15 与 MIG－23 两种飞机在三种飞行状态下的图片。在对目标图像 $f(x,y)$ 抽取了对于平移、比例与旋转变化不敏感的 7 种矩特征后，用 BP 算法进行 3000 次迭

代后达到了网络的收敛，然后用此种训练过的网络对同类飞行的另外一些图片进行识别，达到很高的识别精度，其中对 MIG－23 型飞机达到100%的识别率，对 F－15 型飞机也达到了95.6%。

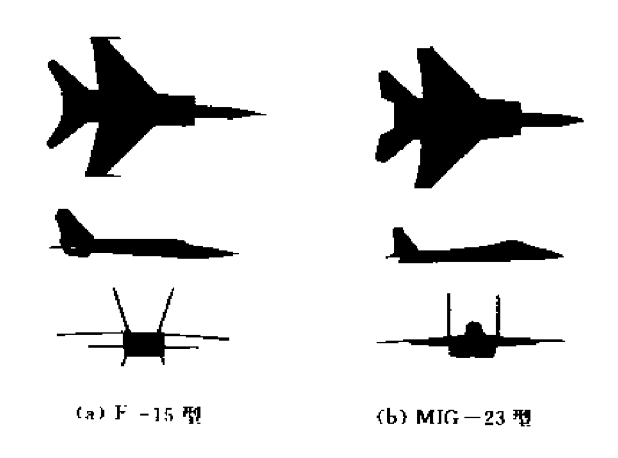


图 9－48 两种飞机在三种状态下的图片

值得指出的是：对于大多数模式识别问题，由于同一类模式样本往往存在相当大的差异，再考虑样本本身的维数很高，所以虽然用神经网络进行直接识别在理论上是可行的，但实际上其时间开销是无法接受的，故通常先进行预处理完成特征抽取，再由神经网络进行识别，不同类的识别对象可抽取不同的特征，但一般说来矩特征特别是高阶的矩特征由于其良好的不变性，更应受到重视。

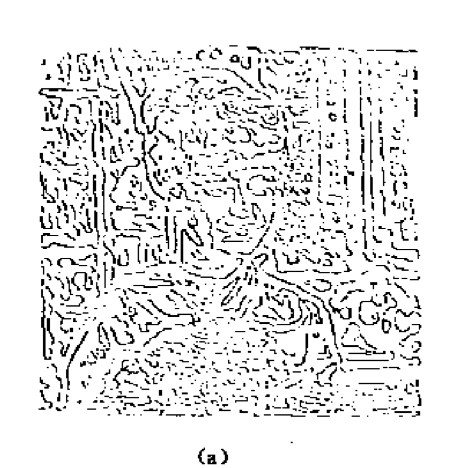
②二维图像的边缘检测[9-3]：

图像的边缘检测是图像识别前必不可少的环节，Marr 和 Hildreth 提出的零交叉边缘检测更是一种十分有效的方法。他们认为①图像强度的突变将在一阶导数中产生一个峰，或等价于二阶导数中产生一个零交叉(Zero-Crosing)；②图像中的强度变化是以不同尺度出现的，故应该用若干大小不同的算子才能取得良好的检测效果。他们认为下述高斯—拉普拉斯滤波器 $\nabla^2 G$ 能满足上述两条原则，其表达式为

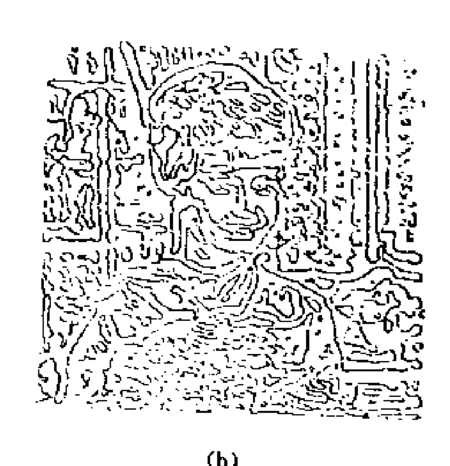
$$\nabla^2 G(x,y)=1/2\pi\sigma^4[2-(x^2+y^2)/\sigma^2]\exp[-(x^2+y^2)/\sigma^2]$$

其中 $\sigma$ 为高斯函数的空间常数，$\sigma$ 值越小，边缘检测算子的敏感程度就越高，图像的细节部分检出越多。

鉴于 BP 模型能完成 $n$ 维空间(输入结点数 $n$)到 $m$ 维空间(输出结点数 $m$)的复杂非线性映射，故它具备 $\nabla^2 G$ 算子的能力。另一方面由于 $\nabla^2 G$ 算子中心激励区的宽度为 $3.6\times 2\sqrt{2}\times\sigma$，故当选取 $\sigma$ 为 1、2、4 时，该激励区的宽度分别为 10、20、40。这一窗口宽度实际上对应于 10×10 或 20×20 或 40×40 个输入结点，从而 BP 网也易于实现 $\sigma$ 的改变。



(a)



(b)

图 9－49 用标准 $\nabla^2 G$ 算子与 BP 网络所作的边缘检测对比

(a)标准 $\nabla^2 G$ 算子　　(b)BP 网络

图 9－49 给出了用两种方法得到的边缘检测结果。(a)图为用标准$\nabla^2 G$算子抽取的边缘，(b)图则为神经网络抽取的结果，它们均是用$\sigma=2$求得的。可以看出，由神经网络可以获得同样良好的结果。而且用神经网络处理，一旦训练完毕，各结点间的互联权就完全确定，在识别时具有很快的速度，若将此网络硬化，则更具有明显的速度优势。

# 习　题

9—1　已知①判别方程为$R=700X_1+80X_2$，式中$X_1$为中等颗粒大小，$X_2$为分类系数。

②训练场中 $A$ 组(滨外沙)向量均值为

$$A=\begin{bmatrix}0.40\\1.20\end{bmatrix}\qquad n_a=40$$

训练场地中 $B$ 组(海滩沙)向量均值为

$$B=\begin{bmatrix}0.30\\0.15\end{bmatrix}\qquad n_b=50$$

求解：(应用判别分析法求)

①$A$、$B$ 二组分组中心的判别值$R_A$、$R_B$。

②判别指数 $R$。

③马氏距离。

④对表中所列样本进行分类

| 类本号 | $X_1$ | $X_2$ |
|---|---|---|
| 1 | 0.51 | 1.14 |
| 2 | 0.46 | 1.31 |
| 3 | 0.33 | 1.40 |
| 4 | 0.34 | 1.19 |
| 5 | 0.55 | 1.17 |

9—2　已知①训练组各分类点处各波段灰度值如题(9－2)表一所示。

②分类树形式如题(9－2)图所示。

③各分类点处的判别函数分别为

$R_{DP1}=32.38ch7/6-22.62ch7/4+5.52ch6/5+16.54ch6/4$

$R_{DP2}=-61.22ch7/6-30.82ch6/4$

$R_{DP3}=-89.47ch7/6+86.69ch7/5+1.00ch7/4-63.01ch6/6$

$R_{DP4}=-61.52ch7/6-29.14ch6/4$

求解：应用序贯分析法将题(9—2)表二中所列象素分出其所属类别。

**题(9—2)表一**

| | $DP_1$ | | $DP_2$ | | $DP_3$ | | $DP_4$ | |
|---|---|---|---|---|---|---|---|---|
| | 水的 $X$ | 其它类 $X$ | 煤的 $X$ | 其它类 $X$ | 水利工程 $X$ | 其它类 $X$ | 沙漠 $X$ | 其它类 $X$ |
| $ch4$ | 11.05 | 29.45 | 11.05 | 29.72 | 25.79 | 29.97 | 34.02 | 27.28 |
| $ch5$ | 7.98 | 43.04 | 10.52 | 43.53 | 37.01 | 43.93 | 52.84 | 38.03 |
| $ch6$ | 2.52 | 46.35 | 9.15 | 46.87 | 56.66 | 46.26 | 56.57 | 39.43 |
| $ch7$ | 0.05 | 36.74 | 3.30 | 37.23 | 52.19 | 36.30 | 44.57 | 30.56 |

# 附录　实用图像处理程序

## 一、3×3 均值滤波器

本程序中，当灰度图像在 f(i,j)周围 8 个象素的平均灰度为 a 时，按下面方法处理象素 f(i,j)：

若　　　　|f(i,j)-a|〈ε，则 f(i,j)←a

若　　　　|f(i,j)-a|≥ε，则不变。

处理后的图像被变换成从 D 到最大灰度共 L 个灰度级。ε 和 L 由用户给定（程序中 ε→eps，L→level，a→averg，f→ii）。

```
/ * 3×3 均值滤波器子程序 filt33
  功能：用 3×3 屏蔽窗口中的 8 近邻均值对 ii[y][x]滤波。中心象素值与其 8 近邻均值的
差超过判定值以上时，该象素值不变，不超过时，用均值置换该象素值。
  输入数组——ii[Row][Col]=输入图像
  输出数组——oi[Row][Col]=输出图像
  输入参数——eps=判定值(0≤eps≤最大灰度值)
  输入参数——level=输出图像的灰度级数(=2,4,8,16,32,64…) * /
filt33(eps,level)
float eps;
int level;
{
    int y,x,gmax=0;
    float averg;
/ * oi[][]清 0 * /
for(y=0;y<Row;y++)
{
  for(x=0;x<Col;x++)
  {
    oi[y][x]=0;
  }
}
/ * 用 3×3 屏蔽窗口的 8 近邻均值进行滤波 * /
  for(y=1;y<Row-1;y++)
  {
    for(x=1;x<Col-1;x++)
```

```
        {
/* 求周围8近邻均值 */

  averg=(float)(ii[y-1][x-1]+ii[y-1][x]+ii[y-1][x+1]
               +ii[y][x-1]+ii[y][x+1]
               +ii[y+1][x-1]+ii[y+1][x]+ii[y+1][x+1])/8.;
/* 中心象素值:ii[y][x]与8近邻均值averg的差超过判定值eps时有无丢失 */
  if(abs(ii[y][x]-averg)>eps)
    oi[y][x]=(int)averg;
  else
    oi[y][x]=ii[y][x];
    }
  }
/* 求oi[][]的最大值gmax */

  for(y=0;y<Row;y++)
    {
    for(x=0;x<Col;x++)
  {
    if(gmax<oi[y][x])
      gmax=oi[y][x];
    }
  }
/* 把oi[][]用灰度级数level整数化以后放入oi[][] */
 for(y=0;y<Row;y++)
   {
   for(x=0;x<Col;x++)
   {
   oi[y][x]=(int)((level-1)*oi[y][x]/(float)gmax+.5);
   write_pixel(320+x,y,oi[y][x]);
   }
 }
}
```

## 二、Robert 微分运算

本程序通过以象素 f(i,j)为中心的 3×3 Robert 算子的作用,生成灰度图像 $f$ 的微分图像 g。用户可选择 3 种灰度梯度定义式,也可以求得梯度方向(−90°～+90°)并放入图像 h(程序中 h→ii)。

```
/ * Robert 微分运算子程序 robert
    功能:生成 ii 的一次 Robert 微分图像 oi
    输入数组——ii[Row][Col]=输入图像(2 值/多值)(保存)
    输出数组——oi[Row][Col]=输出图像(微分图像)
    输出数组——hi[Row][Col]=输出图像(梯度方向:-90 度~+90 度)
    输入参数——kantan=灰度梯度定义式选择
            (0=原定义式,1=绝对值近似式,2=最大值近似式)
    输入参数——hoko=梯度方向计算(0=不计算,1=计算) * /

robert(int kantan,int hoko)
{    int y,x,dl,d2,dd1,dd2;
    int hi[128][128];
    float deg=180/3.14159;

/ * 设定白框 * /

for(y=0;y<Row;y+=Row-1)
  for(x=0;x<Col;x++)
    {  oi[y][x]=200;
    write _ pixel(x+160,y,oi[y][x]);
  }

for(x=0;x<Col;x+=Col-1)
  for(y=0;y<Col;y++)
    {  oi[y][x]=200;
    writ _ pixel (x+160,y,oi[y][x]);
  }

/ * 生成 ii 的 Robert 微分图像 oi * /

  fox(y=1;y<Row-1;y++)
    for(x=1;x<Col-1;x++)
      {
    d1=ii[y][x]-ii[y+1][x+1];
    d2=ii[y][x+1]-ii[y+1][x];
    switch(kantan)
        {

  / * 用原定义式运算 * /
```

```
case 0:
    oi[y][x]=(int)(sqrt((double)d1 * d1+(double)d2 * d2)+.5);
   break;
 /* 用绝对值近似式计算 */
case 1:
    oi[y][x]=abs(d1)+abs(d2);     break;
 /* 用最大值近似式计算 */
case 2:
 dd1=abs(d1);dd2=abs(d2);
   if(dd1>=dd2) oi[y][x]=dd1;
   else oi[y][x]=dd2;
   break;
 }
   oi[y][x]+=64;
   if(oi[y][x]<84) oi[y][x]=0;
   write_pixel(x+160,y,oi[y][x]);

/* 计算梯度方向 */

if(hoko==0)continue;
   if(d2! =0) hi[y][x]=atan((float)d1/d2) * deg;
   eles if(d1)
       hi[y][x]=90;
       else hi[y][x]=-90;

 }
}
```

## 三、图像的相似度计算

本程序计算 2 幅灰度图像 $f_1$,$f_2$ 的相似度 S,S 由下式求出:

$$S=\frac{\sum_y\sum_x f_1(x,y)\cdot f_2(x,y)}{\sqrt{\sum_y\sum_x f_1^2(x,y)}\cdot\sqrt{\sum_y\sum_x f_2^2(x,y)}}$$

相似度 S 越大,表示图像越相似,S=1 时,2 幅图像完全一致(程序中 S→sim)。

```
/* 图像的相似度计算子程序 xsd
  功能:计算 2 幅图像的单纯相似度。
```

```
    输入数组:ii[iy][ix]=输入图像 1
             oi[iy][ix]=输入图像 2
    输出参数:sim=相似度 */

xsd()
{
    int ix,iy;
    int gx,gy;
    int gray1,gray2;
    float sim;
    double norma,normb,c;

/*norma 和 normb 为范数,c 为相关值*/

  norma=0.0,normb=0.0,c=0.0;

  for(iy=0;iy<Row;iy++)
       {
   for(ix=0;ix<Col;ix++)
     {
   gray1=ii[iy][ix];
   gray2=oi[iy][ix];

/*计算范数*/

  norma=norma+(double)gray1*(double)gray1;
    normb=normb+(double)gray2*(double)gray2;

/*计算相关值*/
        c=c+(double)gray1*(double)gray2;
           }
        }
    norma=sqrt(norma);
    normb=sqrt(normb);

/*求得相似度*/
      sim=c/(norma*normb);
    printf("sim=%f",sim);
}
```

## 四、灰度图像的17灰度级表示

程序用图形方式实现灰度图像的伪17灰度级显示。显示中所用的基本象素点用4×4点阵表示，当此点阵中处于发光状态的点数在0～16范围内变化时可得到17灰度级显示，这又被称为Bayer模式。输入图像f可以是任意灰度级的，它被自动地变换为17灰度级。根据显示画面的大小，用于显示f的一个象素的基本象素(4×4点阵)数目可在1,4和16三者中变化。变量不清。

```
  /* 灰度图像的17灰度表示子程序dot17
    功能:把灰度图像变换成17灰度级
    输入数组:ii[im][jm]=输入图像 */
dot17()
{
int gray_matrix[17][4][4]={
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,/* 0level */
    1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,/* 1level */
    1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,/* 2level */
    1,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,/* 3level */
    1,0,1,0,0,0,0,0,1,0,1,0,0,0,0,0,/* 4level */
    1,0,1,0,0,1,0,0,1,0,1,0,0,0,0,0,/* 5level */
    1,0,1,0,0,1,0,0,1,0,1,0,0,1,0,0,/* 6level */
    1,0,1,0,0,1,0,1,1,0,1,0,0,0,0,1,/* 7level */
    1,0,1,0,0,1,0,1,1,0,1,0,0,1,0,1,/* 8level */
    1,1,1,0,0,1,0,1,1,0,1,0,0,1,0,1,/* 9level */
    1,1,1,0,0,1,0,1,1,0,1,1,0,1,0,1,/* 10level */
    1,1,1,1,0,1,0,1,1,0,1,1,0,1,0,1,/* 11level */
    1,1,1,1,1,1,0,1,1,0,1,1,0,1,0,1,/* 12level */
    1,1,1,1,1,1,0,1,1,1,1,1,0,1,0,1,/* 13level */
    1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,1,/* 14level */
    1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,/* 15level */
    1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1/* 16level */
};
   int i,j,im=128,jm=128,thresh=180;
   char sstr[30];
   float fmax;
     printf("Threshhold?");
   i=0;
/* 读入灰度点阵的亮度 */
```

```
do
  {
    sstr[i]=getche();
    if((sstr[i]==27)){sstr[0]=0;sstr[1]=0;break;}
    if((sstr[i]==13))break;
    i++;
}while(1);
    if(sstr[0]!=0)
  {
    int thresh2;
    if(scanf(sstr,"%d",&thresh2))  thresh=thresh2;
}
    color256(0x5d);
    for(i=0;i<17;i++)
    setlut(i,i*3,i*3,i*3);
    setlut(0x01,thresh,thresh,thresh);

/*查找灰度的最大值*/

  fmax=0;
    for(i=0;i<im;i++)
      for(j=0;j<jm;j++)
        fmax=fmax<ii[i][j]?  ii[i][j]:fmax;

/*将原图像变换为17灰度值*/

for(i=0;i<im;i++)
  for(j=0;j<jm;j++)
    ii[i][j]=17.0*(float)ii[i][j]/fmax;
  }
/*灰度图像的17灰度表示*/

display_temp17(int xs,int ys,int Col,int Row,int d,unsigned char *ii)
    {
    int I,J;
    int value;
    for(i=0;i<Row;i++)
      for(j=0;j<Col;j++)
      {
```

```
    for(I=0;I<d;I++)
    for(J=0;J<d;J++)
    {
  value= * (ii+Col * i+j);
  write _pixel(xs+J+j * d,ys+I+i * d,gray _ matrix[value][I][J]);
      }
    }
  }
```

## 五、快速傅里叶变换

void kbfft (pr,pi,n,k,fr,fi,l,il)

1. *参数说明*

pr——双精度实型一维数组，长度为n。当l=0时，存放n个采样输入的实部，返回时存放离散傅里叶变换的模；当l=1时，存放傅里叶变换的n个实部，返回时存放逆傅里叶变换的模。

pi——双精度实型一维数组，长度为n。当l=0时，存放n个采样输入的虚部，返回时存放离散傅里叶变换的幅角；当l=1时，存放傅里叶变换的n个虚部，返回时存放逆傅里叶变换的幅角。其中幅角的单位为度。

n——整型变量。输入的点数。

k——整型变量。满足$n=2^k$。

fr——双精度实型一维数组，长度为n。当l=0时，返回傅里叶变换的实部；当l=1时，返回逆傅里叶变换的实部。

fi——双精度实型一维数组，长度为n。当l=0时，返回傅里叶变换的虚部；当l=1时，返加逆傅里叶变换的虚部。

l——整型变量。若l=0，表示不要求本函数计算傅里叶变换；若l=1，表示不要求本函数计算逆傅里叶变换。

il——整型变量。若il=0，表示不要求本函数计算傅里叶变换或逆变换的模与幅角；若il=1，表示要求本函数计算傅里叶变换或逆变换的模与幅角。

2. *函数程序*

```
#include "math.h"
void kbfft(pr,pi,n,k,fr,fi,l,il)
int n,k,l,il;
double pr[],pi[],fr[],fi[];
{int it,m,is,i,j,nv,10;
 double p,q,s,vr,vi,poddr,poddi;
 for (it=0;it<=n-1;it++)
  {m=it;is=0;
   for (i=0;i<=k-1;i++)
    {j=m/2;is=2 * is+(m-2 * j);m=j;}
```

```
     fr[it]=pr[is];fi[it]=pi[is];
   }
  pr[0]=1.0;pi[0]=0.0;
  p=6.283185306/(1.0*n);
  pr[1]=cos(p);pi[1]=-sin(p);
  if(l! =0) pi[1]=-pi[1];
  for (i=2;i<=n-1;i++)
    {p=pr[i-1]*pr[1];q=pi[i-1]*pi[1];
     s=(pr[i-1]+pi[i-1])*(pr[1]+pi[1]);
     pr[i]=p-q;pi[i]=s-p-q;
    }
  for (it=0;it<=n-2;it=it+2)
    {vr=fr[it];vi=fi[it];
     fr[it]=vr+fr[it+1];fi[it]=vi+fi[it+1];
     fr[it+1]=vr-fr[it+1];fi[it+1]=vi-fi[it+1];
    }
  m=n/2;nv=2;
  for (10=k-2;10>=0;10--)
    {m=m/2;nv=2*nv;
     for (it=0;it<=(m-1)*nv;it=it+nv)
       for (j=0;j<=(nv/2)-1;j++)
         {p=pr[m*j]*fr[it+j+nv/2];
          q=pi[m*j]*fi[it+j+nv/2];
          s=pr[m*j]+pi[m*j];
          s=s*(fr[it+j+nv/2]+fi[it+j+nv/2]);
          poddr=p-q;poddi=s-p-q;
          fr[it+j+nv/2]=fr[it+j]-poddr;
          fi[it+j+nv/2]=fi[it+j]-poddi;
          fr[it+j]=fr[it+j]+poddr;
          fi[it+j]=fi[it+j]+poddi;
    }
  }
  if (1 ! =0)
    for (i=0;i<=n-1;i++)
      {fr[i]=fr[i]/(1.0*n);
       fi[i]=fi[i]/(1.0*n);
    }
  if (il! =0)
    for (i=0;i<=n-1;i++)
```

```
      {pr[i]=sqrt(fr[i] * fr[i]+fi[i] * fi[i]);
       if (fabs(fr[i])<0.000001 * fabs(fi[i]))
          {if ((fi[i] * fr[i])>0) pi[i]=90.0;
           else pi[i]=-90.0;
          }
    else
      pi[i]=atan(fi[i]/fr[i]) * 360.0/6.283185306
    }
  return;
}
```

注:以上程序采用 Turbo C2.0 编写,在 IBM-PC 386/486 微机上皆能运行,其工作环境为 DOS3.0 以上 VGA 显示卡。以上程序中的图像数据大小为 128×128。